

Europe's lesser nuclear powers in flux

By coincidence, last week showed France to be more flexible (and intelligent) than Britain on nuclear weapons, but neither government seems to appreciate the need urgently to put their nuclear forces on a different footing.

EVER since General Paul-Marie Gallois argued in 1956 that France must be a nuclear power capable of inflicting on a potential adversary damage commensurate with the country's value as a prize, France has been the most obdurate of nuclear powers. In the early 1960s, it refused to take part in the moratorium on atmospheric testing, for 20 years it has declined to sign the Nuclear Non-Proliferation Treaty (NPT) and, in Europe since then, has given as one of the reasons (but not the only one) for its studious avoidance of full membership of the North Atlantic Treaty Organization (NATO) its determination to be an independent nuclear power. Britain, by comparison, has been a goody two-shoes — a founding member of the Partial Test-Ban Treaty, of the NPT and all that. Are the roles now reversed?

Ten days ago, French President François Mitterrand said that France would be willing to regard its nuclear weapons as in trust for the defence of Europe. By itself, of course, that remark means very little. At this stage, there is no suggestion that France will cease to be an independent nuclear power, while there is nothing to suggest that Mitterrand had asked the other states of Europe whether they would welcome his suggestion. But his offer is a sign of flexibility not seen for decades, and an opening for discussion.

Immobility

By contrast, little seems to have changed in Britain. In an arid debate in the House of Commons last week (14 January), the government rehearsed its long-standing position that, while eastern Europe and Russia remain in turmoil on defence, past policies must simply be extrapolated into the future. Among other things, that entails the building of a fourth nuclear submarine to carry Trident nuclear missiles. The Labour opposition party, on the other hand, would put its weight behind further negotiations on strategic and conventional arms control in Europe, and is not persuaded of the need for a fourth nuclear submarine. Even allowing for the paralysis of intellectual imagination by the impending election, the arguments of both sides are a poor response to an historic opportunity — that of exorcising the nuclear weapons from central Europe's strategic balance.

But the upheaval in the East is not only an opportunity, but a source of great uncertainty. The arms control agreements concluded during Mr Mikhail Gorbachev's

spell at the Kremlin were far-reaching, but there is now only the sketchiest assurance that they will be enforced (which will create problems in the US Senate when ratification comes up). Certainly this is not the time to be launching a new set of formal arms control negotiations, which will have to wait until the dust has settled.

Opportunity

Moreover, while the new republics (notably Russia and the Ukraine) are still squabbling about the division of the military spoils of the dismembered Soviet Union, there are no means by which they or anybody else can tell whether they will in future be middle-sized nuclear powers of some kind. The signs of flexibility in France, however wraithlike, may encourage the new republics to avoid dependence on nuclear weapons in their military defence; the rigidity of the British position (on both sides of the parliamentary aisle) will work in the other direction.

Yet neither Britain nor France can hope that its strategic nuclear forces can now be left unchanged. Indeed, it has for many years been understood (and acknowledged by at least British governments) that substantial cuts in superpower strategic arms would compel some kind of change. Mitterrand's vague promise suggests that he also sees the writing on the wall. So should not the French and British governments, both members of what Brussels wants to call the European Union, be coordinating their policies on nuclear weapons, and at the same time talking to the leaders of the new republics? Such hopes are not new, but have usually foundered on the traditional French distrust of the United States as the linchpin of NATO. As things are, coordination within the framework of Western Europe's own military alliance, called Western European Union, is the obvious course.

Despite the election in the offing, the British government should be getting on with that. (There might even be a few votes in a show of flexibility.) And both governments have everything to gain from taking up the question of what the new republics intend on nuclear defence, without waiting for the remarkable meeting planned for the end of this month between the permanent members of the UN Security Council and Mr Boris Yeltsin, the President of Russia. It appears quickly to have been forgotten that the remarkable sequence of arms control agreements reached in the 1980s rested firmly on the endless arguments on strategic and tactical strategy in which the old Soviet



Union had been engaged long before Gorbachev's time. Now, there is even more tortuous discussion to be had. It is in everybody's interests that the republics should be tempted towards modesty in their reliance on nuclear weapons; countervailing self-restraint in Western Europe can only help.

And the timing is critical. In three years, the NPT will lapse unless its present members agree to soldier on. With all the shocks that system has suffered in the past few years, notably in Iraq, it is not too soon to begin working on the amendments to the treaty required for its perpetuation. Apart from the technical work needed on the supply of nuclear weapons, there are the strongest reasons to believe that a willingness by Britain and France to displace nuclear weapons from the heart of their national (as distinct from regional) defence would go a long way to cool tempers at the conference due in 1995. □

Russian roulette?

Some parts of the Russian research establishment seem bent on firing those who spend too much time elsewhere.

GENERAL sympathy with the plight of the research community in what used to be the Soviet Union may not last as long as it will be needed, at least if some Russian reactions to public bankruptcy become widespread. There are two spheres in which money problems bite. Hard currency, called 'currency' in the vernacular, has virtually disappeared from the research enterprise, with the result that reagents and journals previously bought from outside have simply become unobtainable. And, now that the printing presses have been halted (or are manufacturing rubles less quickly), research institutes are unable adequately to support their own members. For the past three years, most of them have been free to adjust individual salaries by the use of a proportion of whatever income they have from outside. Now, some of them appear to be turning to intellectual cannibalism of a disturbing kind.

Researchers able to make what are called 'business trips' abroad have always been enviously regarded in the Soviet Union. In the bad old days, the frequency and duration of such visits was more a mark of indulgence and patronage than merit, but that has mercifully changed in the past few years. But now there is a risk that seriously intended and welcome visits to laboratories elsewhere will dry up. That, at least, may be the consequence of the decision by some institutes of the Russian Academy of Sciences at Novosibirsk that researchers who spend three months or more at an overseas laboratory may be required to pay for the privilege with their jobs. They may be sacked.

A more efficient way to encourage mediocrity could hardly be devised. The people who trickle overseas in these competitive times are the brightest and the best; ideally they are also young. It is in everybody's interest, and not least that of the Russian research enterprise itself, that they should be welcomed back to posts enjoying

more, and not less, of what little security the new republics can offer. It is to be hoped that the Novosibirsk recipe will not be followed generally — and that it will be abandoned even there. □

Psychics in court

Mr Uri Geller, the spoon-bender (who is still among us, but less obtrusively than in the past) seems to have stumbled into an awkward dilemma. As part of the trial of a legal action for libel against Mr James ('the Amazing') Randi, a Florida Court has now decided that Geller must perform a demonstration of his psychic powers. Only then, the court seems to have argued, will it be possible fairly to assess Randi's disputed allegations that Geller's powers are not out of the ordinary. As part of the discovery process, the court requires that Geller's demonstration should be recorded by video camera.

The occasion should have a significance extending beyond the suit being tried in Florida. Geller is not, of course, the only one to claim psychic powers. There is an army of astrologers and a battalion of faith-healers whose claims are similarly disputed by rational people. Sadly, their claims on public attention are too often untested in any objective way. Indeed, a degree of mystery seems inseparable from their success; only a few weeks ago, Geller told the Israeli newspaper *Yediot Achronot*, "I prefer that no one should ever know if I am real or not. It's better that there should be some controversy surrounding me." The truth of that assertion is easily understood. Whatever the validity of the psychics' claims, their use of them is usually inseparable from show-business of a kind. Unintentionally, the Florida courts have now struck a blow for objectivity by commanding a test of Geller's psychic powers. In the public interest, it is earnestly to be hoped that he takes part in the demonstration. □

Yale retrenches

YALE University has taken a bold step in its announced intention to face fiscal reality by cutting back some academic departments and eliminating others (see p. 288). For years US universities, whose science departments were generously funded by government grants that also provided substantial overhead costs to the general university administration, have lived as if money were no object in the pursuit of education and knowledge. Now reality has struck. Yale is not the first university to talk about eliminating whole departments (in this case linguistics, operations research and an institute for social and policy studies), but it is the most prominent. So far. Yale is at the forefront in recognizing the urgency of making hard choices and it is to be commended for doing so. It is to be hoped that intramural politics will not prevent Yale from doing what is, in principle, necessary and altogether sensible. □

CERN told to pay \$70 million to engineering consortium

- Money will be out of pocket
- Current programme to continue

London

CERN, the European particle physics laboratory in Geneva, has suddenly found itself more than \$70 million out of pocket. After six years of legal wrangling, a Swiss court has ruled that CERN should have paid an extra SFr60 million to the consortium of civil engineering companies that built the 27-km tunnel housing the laboratory's Large Electron-Positron (LEP) collider on top of the SFr292 million contract agreed before the EUROLEP consortium began tunnelling. With interest, the total bill runs to a cool SFr100 million (about \$70 million). As the CERN member states are unlikely to provide that much money in addition to their existing CERN subscriptions, the court decision will delay large new projects — including the Large Hadron Collider (LHC), which European physicists hope will be the first machine to detect the elusive Higgs boson.

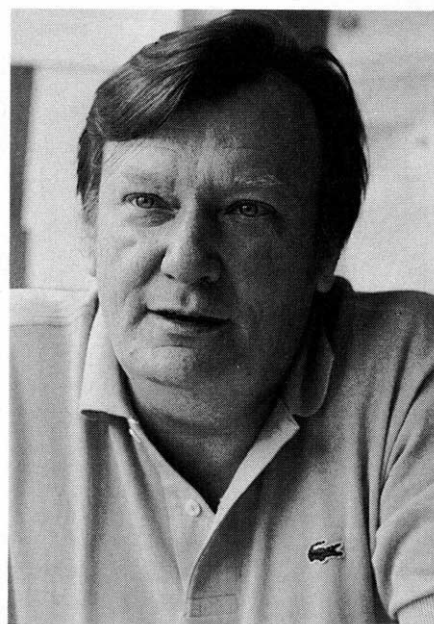
When drilling the LEP tunnel in the mid-1980s, EUROLEP found the tunnelling conditions more difficult than expected, and had to bring in extra manpower and equipment to complete the tunnel on the tight schedule required by CERN. (At the time, the laboratory was competing with US physicists at the Stanford Linear Accelerator Center to be the first to generate large numbers of Z^0 particles; the detection of Z^0 s at CERN and Stanford subsequently showed that there are only three families of elementary particles (see *Nature*, 341, 555; 1989)). EUROLEP argued that CERN should foot the inflated bill for building the tunnel.

The court's decision is a hard hit to take," says CERN director-general Carlo Rubbia, but he is relieved that EUROLEP's claims were not accepted in full: the consortium had been pressing for more than SFr600 million (about \$420 million), including interest, on top of the original contract. "That would have been the death of our laboratory," says Rubbia.

CERN's governing council, which meets next in March, now has to decide how to find the money. CERN cannot appeal the decision, and the bill must be paid quickly, to avoid further interest charges. Sandy Donnachie, from the University of Manchester, and a British delegate to the CERN council, expects delegates to put pressure on France to provide at least some of the money. CERN is a valuable source of employment and income for the surrounding area, both in Switzerland and over the border in France. Switzerland does pay an extra supplement

on top of its CERN subscription in recognition of this, but France has always declined to do so. In any case, says Donnachie, "the bulk of the companies involved [in EUROLEP] are French".

Rubbia accepts that he will probably have to find "a significant fraction" of the SFr100 million from CERN's existing budget, but says he has no intention of



CERN director Carlo Rubbia

closing any of the machines now operating at the laboratory. Instead, new projects — a planned upgrade of LEP, and the construction of its successor, the LHC — may have to be postponed. But Rubbia is not too concerned, because he says the delays can be measured in months, rather than years. With the United States having failed to secure a Japanese contribution for the LHC's rival, the Superconducting Super Collider, "CERN has so little competition today, that it's a shock that we might be able to absorb," he says.

Nevertheless, the March CERN council meeting will be a lively one, Donnachie predicts. "One could criticise CERN to some extent for not setting up a contingency fund" to pay EUROLEP, he says. But Rubbia says that CERN's priority has been repaying the money it borrowed to build LEP. Debt repayments have been running at around SFr85 million (about \$60 million) a year, and were due to finish in 1993. The new SFr100 million bill may now force the loans to be extended, however.

Peter Aldhous

DIETARY SUPPLEMENTS

'IQ pills' land in trouble

London

HEALTHCRAFTS, the company that last year launched a vitamin and mineral dietary supplement on the back of a controversial research claim that the formulation could boost children's intelligence (see *Nature* 350, 5; 1991), faces charges under British trading law. Trading standards officers in Shropshire, in the English midlands, accuse Healthcrafts of making misleading statements in advertising and packaging their product Vitachieve. The research on which the claims for Vitachieve were based was conducted by Stephen Schoenthaler, from California State University, and his colleagues, who last March reported an increase in nonverbal IQ scores averaging 3.7 points among children given the supplement for 13 weeks. At the time, the research attracted great controversy, not least because of Healthcrafts' decision to begin marketing Vitachieve on the basis of the results, with the endorsement of the Dietary Research Foundation — the charitable trust that funded the work.

The case is likely to focus on the challenges made against the findings by others in the field (see *Nature* 350, 13; 1991), and on the fact that Schoenthaler and his colleagues say that the supplement will, in any case, benefit only a minority of children with poorer nutrition (see *Nature* 351, 263; 1991). Healthcrafts' advertisement said that the average IQ score of children taking the supplement had increased significantly, and was headed: "Extraordinary news for parents."

Charges are also being brought against a second company, Larkhall Natural Health Ltd, over the marketing of a similar supplement, Tandem Ideal Quota. The claims made for Tandem IQ refer to an earlier research project, conducted by David Benton from University College, Swansea, and Marilyn Roberts, a Wrexham schoolteacher. The claims, published in *The Lancet* (i,140; 1988), also concluded that dietary supplements can influence IQ.

A spokeswoman for Healthcrafts says that the company cannot comment on the charges for the time being, but Robert Woodward, managing director of Larkhall, disputes the accusation that Tandem IQ is marketed using misleading claims. He says that the product's packaging states explicitly that the children who benefited most in Benton and Roberts' experiment were those with the poorest diets. He intends to fight the case. But he adds that last year's controversy surrounding the launch of Healthcrafts' product has made the British public "extremely cynical" about the influence of dietary supplements on IQ. "It really killed the market," he says.

Peter Aldhous

Daniel Zagury cleared of misconduct charges

Washington

ALLEGATIONS that Daniel Zagury violated medical codes of ethics in experimental trials of an AIDS vaccine are unfounded, according to French medical officials who investigated the charges that were first raised by a reporter from the *Chicago Tribune*. Three of Zagury's patients, each in the advanced stages of AIDS, died from a reaction to the experimental vaccine.

The French Conseil Régional de l'Ordre National des Médecins not only concluded that Zagury and his colleagues, acted ethically in their experiments at the Saint Antoine Hospital in Paris, but also pointed out that all medical trials necessarily include some risk and that by participating in the vaccine trials the patients did not face a risk "out of proportion to [their] state of health and diagnosis". Each had T4 lymphocyte counts of less than 50/mm, which usually puts AIDS patients close to death.

Zagury's judges went so far as to say that with AIDS, "research is a duty" and that "it is legitimate for [Zagury] to follow this work on active immunology in order to determine whether it offers a useful treatment for this disease."

Since 1987 Zagury has been working on what he calls "active immunotherapy" for AIDS patients — a vaccine that uses the patients' own cells, infected *in vitro* with inactivated recombinant vaccinia virus engineered to express human immune deficiency virus (HIV) antigens. In an effort to boost AIDS immunity in the three patients who were subsequently found to have died of vaccinia necrosis, Zagury administered the vaccine by three routes — intravenous, intramuscular and subcutaneous injection.

Zagury says that he did not immediately suspect vaccinia necrosis as the cause of death in his first two patients, mainly because he (and others) had considerable experience with inactivated vaccinia and, even though it can be hazardous to immunocompromised individuals, "at first we did not think that was it". Once another researcher, using monoclonal antibodies to search for vaccinia antigen, found it in the keratinocytes of the deceased, Zagury says he was convinced that their deaths were due to vaccinia. "At that point, we stopped the subcutaneous injections altogether — in all patients in the trial," he says.

In its report exonerating Zagury of ethical misconduct, the council notes that he sought approval from the hospital ethics committee in 1987, before such approval was required by law, and again in 1990, and concludes from this that Zagury and

his colleagues were acting openly and in compliance with ethical cannons. Furthermore, the patients gave written

Associated Press

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Zagury vindicated.

informed consent for the studies and were covered by an insurance policy that covers experimental studies.

It is no surprise that Zagury is relieved by the exoneration which, he says, is important not only to him but also to "my family's honor and to medicine".

Barbara J. Culliton

MEDICAL RESEARCH

Will EC policy boost European health studies?

London

THE European Commission may in future become an important player in medical research funding, under the new European Communities' (EC) treaties agreed at the Maastricht summit of European leaders in December, which commit the EC states to greater economic and political unity. Buried in the reams of paper now being studied by EC lawyers is a chapter on public health, which says that the Commission may support research into the most important health problems in Europe.

This is the first time that the Commission has been given an explicit mandate to promote medical research aimed at problems of public health. If interpreted liberally, the wording could lead to a significant increase in the Commission's spending on medical research. The current five-year EC medical

Call for UK gene therapy

London

BRITAIN needs a new expert panel to assess proposals for human gene therapy experiments, according to an independent committee chaired by the lawyer Sir Cecil Clothier, which was asked by the UK Department of Health to consider the ethical questions associated with gene therapy. As expected, the report concludes that somatic gene therapy poses no special ethical problems, but recommends a moratorium on the manipulation of germ line cells (see *Nature* 355, 190; 16 January 1991). Somatic cell gene therapy should only be used to alleviate genetic disease, not to change normal human characteristics, the report recommends.

In addition to scrutiny by local research ethics committees at hospitals where gene therapy is planned, the Clothier committee says that all proposals should be vetted by a national body (A similar procedure is followed in the United States for all government-funded research.) This should contain scientific and medical experts, able to advise on the conduct of the research, and lay representatives. The Department of Health intends to consult with interested bodies for the next four months, before acting on the report's recommendations. Until a new national committee is set up, the Clothier committee will stay in place, to review any gene therapy research proposals that may come forward. **P.A.**

research programme, just getting under way, amounts to only 131 million ECU (approximately \$92 million).

Commission officials say it is too early to predict the precise effect that the new profile given to public health will have on EC research spending. The Maastricht agreement has yet to be ratified by the EC member states (something that could take a year or more), after which a high-level committee set up by social affairs commissioner Vasso Papandreou will decide which diseases to target with new research programmes.

The text of the agreement reached at Maastricht mentions drug dependency as one example of priority, but it is expected that the prevention of cancer and cardiovascular disease — Europe's major killers — will be priority areas for research as well.

Peter Aldous

ARTIFICIAL ENVIRONMENT

Japan builds mini earth

TOKYO Japan's Science and Technology Agency (STA) plans to create a "mini Earth" within a large sealed concrete and glass building in the little village of Rokkasho on the remote northern tip of mainland Japan to study the global environment, to develop facilities that can create artificial environments for man to live on the moon, and to determine the effects that radiation leaks from nuclear power accidents will have on the natural environment.

The village of Rokkasho is famous for a huge \$8,000-million facility for recycling nuclear fuel that is being built there much to the concern of some local people (see *Nature*, **345**, 285, 24 May 1990). The village of a few thousand farmers and fishermen has not until now been seen as a centre of global environment research. But STA thinks it soon will be.

Last week, the agency announced plans to build the mini Earth, or biosphere as it is otherwise called, by 1995 with a total investment of 4,000 million yen (\$32 million). About 700 million yen (\$5.6 million) has been set aside for the project in the fiscal 1992 budget announced earlier this month to buy land and to design the facility, which will be next door to the nuclear fuel recycling plant.

The idea is to build a sealed concrete and glass structure with a floor space of about 1000 square meters in which plants and animals will be kept in an artificially created and controlled atmosphere. Experiments will be carried out to try and understand the carbon dioxide cycle by constantly monitoring carbon dioxide and oxygen levels in the biosphere with computers.

In other experiments, trace amounts of radioactive substances will be released into the biosphere to see how they spread among living organisms. And by building the facility, planners hope to learn how to create artificial environments for man to live on the moon or under the sea.

The National Space Development Agency and the national radiological research institute will participate in the research, which will be run by a little-known new foundation in Rokkasho called the Environment, Science and Technology Research Institute. Agency officials admit that the institute, which only opened in December 1990, has very few researchers. But they hope to have many more in the future.

Whether top-class researchers on the global environment will want to go to one of the most remote and bleakest parts of Japan remains to be seen.

Observers say the prime purpose of the facility seems to be to appease local fears about the nuclear fuel plant and to convince the local people that everything that can be

done to ensure safety of the plant is being done. In addition to the biosphere experiments, the institute will carry out experiments to determine the effects of low-level radiation on 20,000 mice.

David Swinbanks

DRUG DEVELOPMENT

Drugs tested for women

Washington

DESPITE the ethical, legal and safety problems of testing new drugs in women, particularly those of child-bearing age, a recent survey* by the Pharmaceutical Manufacturers Association (PMA) has identified 263 drugs that are being developed for use in women. The medicines, which are in human clinical trials or awaiting approval by the US Food and Drug Administration, are for diseases and conditions that either exclusively or mostly affect women, or are among the top ten causes of death in women.

According to the survey, the top three research categories are cancer, obstetric and gynaecological diseases, and cardiovascular/cerebrovascular diseases. The three leading causes of death in women — heart disease, cancer and cerebrovascular disease — are garnering 40 per cent of the research effort, the survey finds. In addition, the survey points out that more drugs, 37 in total, are being developed for the treatment of breast cancer — the second most common cause of cancer death in women — than for any other disease affecting women.

In a separate but related survey, pharmaceutical companies were asked whether they routinely test medicines in women and monitor data for differences in gender. Of the 33 companies that responded, 94 per cent said that they always collect data on the gender of participants in a clinical trial, six per cent said usually. When asked if they deliberately recruit representative numbers of women for clinical trials, 76 per cent of companies said yes.

Although the survey indicates that pharmaceutical companies and the biomedical research community are paying closer attention to the special medical needs of women, it urges that more progress must be made in understanding why women are more prone to certain diseases and why certain medicines work differently in men and in women.

Diane Gershon

* New Medicines in Development for Women (Pharmaceutical Manufacturers Association, Washington, DC, 1991).

DRUG POLICY

FDA panel backs Interleukin-2

Washington

AN advisory panel to the US Food and Drug Administration (FDA) has advised FDA commissioner David Kessler to approve the use of interleukin-2 for the treatment of patients with terminal kidney cancer. The panel's recommendation for renal carcinoma comes despite the fact that interleukin-2 is known to be toxic to the heart and other organs. The recommendation is consistent with a current trend to let patients decide whether to submit to dangerous therapy when no effective alternatives are available.

In the United States alone, an estimated 25,000 people are diagnosed with kidney cancer every year. In clinical trials, kidney tumors shrunk in 15 per cent of 255 patients, though none was cured.

The panel's approval is good news for the Cetus Corporation of Emeryville, California which spent more than \$100 million on the development of this drug. It is also good news for Chiron, which merged with Cetus last year. The drug, which is expected to get final FDA approval, will sell under the trade name Proleukin.

B.J.C.

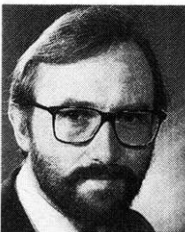
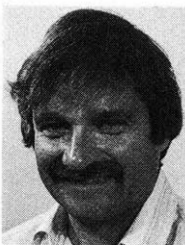
RESEARCH PRIZE

Three share Jeantet award in medicine

London

EUROPE'S largest award for medical research, the SFr2 million Louis Jeantet prize, has this year been shared by Paul Nurse and Alain Townsend, from the University of Oxford, and by Christiane Nüsslein-Volhard, from the University of Tübingen in Germany. Nurse is recognized for his discovery of the molecular components of the yeast cell cycle, Townsend for his research on recognition of viral antigens by T lymphocytes, and Nüsslein-Volhard for her developmental genetics studies in *Drosophila* and other species. All of them say they plan to use the money to expand their research groups.

P.A.



Nurse (top), Townsend (middle), and Nüsslein-Volhard.

ERS-1 data still in the works

London

MANY of the researchers chosen as principal investigators (PIs) for the European Space Agency (ESA)'s ERS-1 remote sensing satellite are getting impatient. More than five months after the satellite's launch, PIs are still waiting to begin receiving data on a regular basis.

ERS-1 is considered by many to be the most advanced remote sensing satellite yet flown, and is expected to provide a huge volume of oceanographic data, which will help to improve current climate prediction models. The PIs, who number more than 270, were hoping to begin receiving regular instalments of ERS-1 data late last year. But most have so far only been given a tantalizing taste of ERS-1's potential, in the form of single batches of test data.

PIs and officials from the ESA member states alike are confused about the reasons for the delay. We are beginning to get concerned," says Trevor Guymer, a PI from the James Rennell Centre for Ocean Circulation in Southampton. David Williams, an official responsible for remote sensing at the British National Space Centre, says that the issue was raised by many national officials at a meeting in Paris earlier this month. Everybody had the same concern," he says: "Why have the data not gone open?"

ERS-1 carries a range of microwave remote sensing equipment which, unlike instruments operating in the visual spectrum, can detect the Earth's surface even through heavy cloud cover. Its centrepiece is the Active Microwave Instrument (AMI), which can operate as a Synthetic Aperture Radar (SAR), producing high quality radar images of a 100-km-wide strip of the Earth's surface. These images will be used in land use surveys, to monitor the movement of ice or oil slicks at sea, and may even allow oceanographers to locate ocean fronts, where bodies of warm and cooler water meet. Alternatively, the AMI can be switched to work as a scatterometer, measuring the backscatter from radar beams angled at 45 degrees to the sea surface—data that can be used to calculate wind speed and direction. ERS-1 also carries the Along Track Scanning Radiometer, which will measure the water content of the atmosphere and the sea surface temperature, and a radar altimeter, to measure the height of ice sheets and variations in sea level.

Some ERS-1 data are intended to be used within a few hours of collection, mostly by Europe's meteorological agencies. Distribution of these fast-delivery products seems to have begun without problems. (The European Centre for Medium-Range Weather Forecasts in

Berkshire, for example, began receiving fast-delivery products in August last year, although staff at the centre are still experimenting with their use, and are not yet ready to use ERS-1 data in operational weather forecasting.) But most of the PIs are interested in so called 'off-line' data from the satellite. These are high quality data products, including SAR images and oceanographic measurements calculated from raw ERS-1 data.

"Maybe there has been some delay" in supplying PIs with off-line data, concedes Stefano Bruzzi, ESA's deputy mission manager for ERS-1. But he says that the agreement between ESA and the PIs was that each PI must return a report on the use of a test data set, before receiving regular instalments of data. So far, ESA has received very few of these reports, says Bruzzi. But Meric Srocosz, from the James Rennell Centre, says that many of the test data sets were not sent out until mid-December, and there simply has not been time to complete the reports. ESA originally said that test data would be issued by 15 October, he says. "It's tricky to send a report on data we hadn't received."

Other researchers are more sympathetic towards ESA. Stephen Briggs, from the Institute of Terrestrial Ecology at Monks Wood in Cambridgeshire, acknowledges his colleagues' concerns, but believes the problem will soon be resolved. ESA originally estimated that it would take three months to calibrate the instruments on board ERS-1, and to validate the algorithms used to calculate the off-line

data products, by comparing ERS-1 data with 'real' measurements taken from ships, buoys and aircraft. In the event, this process took some five months, says Briggs.

"ESA have set themselves very ambitious targets," agrees David Llewellyn-Jones, from the Rutherford Appleton Laboratory in Oxfordshire, who heads the team which built the Along Track Scanning Radiometer. He believes that ESA and the PIs simply underestimated the logistics of handling the data. The distribution network is complex: raw data beamed down to ground stations throughout the world are first transmitted to ESA's data handling centre at Frascati in Italy. To produce the off-line data products, data are then relayed to three Processing and Archival Facilities, in Britain, France and Germany (a fourth, in Italy, is not yet fully operational). With ERS-1 churning out prodigious quantities of information MD the SAR produces data at a faster rate than any instrument flown previously on a civilian satellite MD the distribution network always faced a formidable task.

Despite the disappointment over the delays, most PIs agree that, once the data start flowing, ERS-1 should give ESA a world lead in microwave remote sensing. The field was pioneered by the US National Aeronautics and Space Administration (NASA), which flew a SAR instrument on board the short-lived Seasat mission in 1978 (the satellite failed after only three months in orbit, but its data are still being analysed by oceanographers). NASA has since failed to consolidate on this early lead, carrying SAR instruments only on two brief space shuttle flights.

Peter Aldhous

UNIVERSITY POLICY

Yale to cut departments

A YALE University faculty committee, in what may be the wave of a painful future, last week concluded that, in order to maintain academic strength in most disciplines, Yale should eliminate some departments and consolidate others during the next few years.

The 12-member faculty committee, which spent the past year examining the university against what is called "the financial situation at Yale", has called for the elimination of the departments of linguistics and operations research. Neither department is sufficiently strong and neither attracts enough students to justify continued existence as a full department, the committee says.

It also argued for the elimination of Yale's Institution for Social and Policy Studies on grounds that its activities "have not been commensurate with the resources devoted" to it. A reduction in the size of the sociology department was also

proposed.

Engineering research also made the list of departments that do not quite meet the test in a financially tight era. The recommendation is to merge chemical engineering, electrical engineering and mechanical engineering into a single Engineering department, with a somewhat reduced number of faculty positions. The merger of physics and applied physics is also called for. Cumulatively, these consolidations would result in a reduction of 49 faculty positions if the Yale's governing board accepts the faculty committee's recommendations.

In conducting its analysis, the faculty committee was asked to consider the gradual elimination or consolidation of departments, rather than a percentage cut in funding for departments across the university. Final action on the restructuring plan is expected sometime within the next year.

Barbara J. Culliton

The Danube: Dams over troubled waters

London

HUNGARY is threatening to cut off river, rail and road traffic to Czechoslovakia in retaliation for the imminent resumption of construction of the controversial Gabčíkovo–Nagymaros hydroelectric project on the Danube River. The emotionally charged project provoked heated new demonstrations last year in Prague and Vienna as well as a dozen West European cities.

Hungary has abandoned its part of the hydroelectric construction programme at Nagymaros, in the scenic Danube Bend north of Budapest, for environmental reasons. But Slovakia, the smaller of the two states comprising Czechoslovakia, has decided that its own construction work already carried out on the Gabčíkovo dam south of Bratislava has dealt such irreparable damage to the environment that it should now be completed as the lesser of two evils.

Austria is seeking compensation for its original \$640 million investment in the 15-year-old joint programme which should have been repaid in hydro-power by now. Hungary and Czechoslovakia have each invested about \$500 million, but most of the construction has taken place in Czechoslovakia where officials insist that the unfinished project cannot just be abandoned.

Environmentalists in all three countries have opposed the project from its inception, arguing that the dams would change the flow of ground water, pollute drinking water, damage agriculture and forestry and dry up the river's vast inland delta which constitutes a unique wetland rich in wildlife.

The two dams would need to operate in conjunction to reap the maximum potential power yield of the river. The Gabčíkovo project, which is now expected to be completed "provisionally" this year, should produce in about three years 180 MW instead of the originally planned 780 MW of electric power.

The original plans of the joint project would have involved a re-drawing of the border between the two countries, which cannot now be completed without Hungary's permission. The border is the river itself which was to have had a weir built across it to divert part of the flow. Hungary has just demonstrated its determination to wind up the project by inviting tenders for demolition of the unfinished barrage in the Danube Bend and restoring that section of the river to its former state.

Not so Czechoslovakia. Under a compromise plan known as the 'C' variant,

a 17 km canal already completed in Slovakia will now be extended by another 9 km upstream where both sides of the river are on Czechoslovak territory and water management can therefore be treated as an internal matter.

If Hungary retaliates by blocking the trade arteries of the landlocked region, there could be far-reaching effects. Public opinion in Hungary would certainly support such a drastic measure. So would important segments of population in



Danube dams spur controversy.

Czechoslovakia — particularly the large Hungarian ethnic minority in Slovakia. The influential global environment protection movement would also be sympathetic. But a disruption of goods transport would damage business and commerce in Central Europe just as the region is trying to pick its way out of the economic ruins of communism by changing to a market economy.

The Swiss-based World Wide Fund for Nature, which has taken a special interest in the controversy, says the construction of the dam in Slovakia has already required the razing of 100 sq.km of forest and farmland and the devastation of a 40 sq.km area which has been covered with concrete.

The mayors of 80 towns adjacent to the project area are supporting a petition seeking a moratorium on new construction. They are campaigning for the establishment of an international scientific commission to look into the broad economic as well as environmental consequences of the project.

But not all news of the Danube is gloomy. All the countries of the river basin are involved in a programme launched late last year under the auspices of the United Nations and the European Community to clean up the common environment. The \$35 million first phase of the programme will bring together many scientific research and training institutions to improve water pollution monitoring and data management among other things.

Thomas Land

Biotech lobby pressures EC

London

EUROPE'S biotechnology companies are stepping up their efforts to combat what they see as the European Communities' (EC) muddled policies towards the industry. The Senior Advisory Group on Biotechnology (SAGB), set up in 1989 by nine of Europe's largest chemicals companies (including ICI, Hoffmann-La Roche and Ciba Geigy), is expanding to include 28 of Europe's leading biotechnology concerns. The move recognizes the fact that the policies of the EC, not those of individual governments, now have the biggest influence over the European biotechnology industry.

The advisory group has consistently argued that existing product-safety legislation is sufficient to regulate the industry, and that additional EC controls over genetic engineering will make it difficult for Europe's biotechnology companies to compete with US and Japanese competitors (see *Nature* 343, 501; 1990). A year ago, SAGB called projects being supported by the EC "too limited in both scope and scale to have a lasting impact and give a clear European profile".

Brian Ager, from SAGB's office in Brussels, says that the European Commission's stated policy on biotechnology, released last year, recognized this concern, but he believes that continued pressure is needed to ensure that this policy document is acted upon. "We still haven't got it right in Europe," he says.

The expansion of the advisory group follows the establishment of another industry lobby group, the European Secretariat of National BioIndustry Associations, which also plans to open a Brussels office. Its first concern is likely to be a directive on the transport of genetically engineered organisms, now being drafted by European Commission transport officials, which biotechnology companies fear will be far more restrictive than anything planned in the rest of the developed world. The advisory group hopes the EC will treat genetically engineered products according to their inherent characteristics, not method of manufacture.

A major priority for both lobby groups is the adoption of a planned EC directive on the patenting of biotechnological inventions, first launched in 1988, but which has become stalled in the European Parliament amid controversy surrounding a proposal to allow the patenting of transgenic animals.

Peter Aldhous

Air pollution and health

SIR — In the light of recent discussions of the effects of air pollution on human health, and of the need for more detailed examination of the effects on those at risk, we write to inform your readers of a joint project by the Italian Research Council (NRC) and the Italian Electric Power Authority, funded by L35,000 million over three years. We believe the epidemiological study that is a major part of this programme has several important and distinctive features.

Two general population samples have been enrolled for longitudinal studies, one in the rural area of the Po River delta, where there is a large thermoelectric plant, and one in the urban and suburban area of Pisa in central Italy, which is polluted mainly by traffic exhausts.

The studies we plan to carry out will allow a measure of both the classical differences of prevalence and incidence rates of respiratory symptoms and diseases and of the cross-sectional and longitudinal differences of lung function after allowing for confounders such as smoking and socioeconomic status.

We also plan to study the distribution of nonspecific bronchial hyperreactivity (methacholine challenge) and the distribution of skin test reactivity and IgE levels in the two populations. We plan to use haemoglobin adducts to benzene and benzopyrene as biological markers of exposure to vehicle exhausts and nitrous compounds to mark exposure to NO_x . We shall score DNA lesions (sister chromatid exchange, micronuclei and so on) in cultures of lymphocytes from the urban subjects and from a subsample of the rural population.

There will also be a series of nested case-control studies of subgroups selected for specific characteristics (for example, non-smokers in the urban areas with the highest and lowest levels of pollution). Serum antibodies to DNA adducts (benzopyrene) will be measured, as will be the presence of DNA adducts in lymphocytes and possibly in alveolar macrophages collected from volunteers from the same subsamples.

The project follows a complex multidisciplinary design. Experts in epidemiology, lung diseases, immunology, allergy, biochemistry, cytogenetics, environmental chemistry and computer science are involved. The sample of the urban-suburban area of Pisa consists of 3,800 people between the ages 8 and 70 selected on the basis of family clusters taken from the most recent census. We have taken account of socioeconomic status, age and sex so as to arrive at a sample representative of the general population.

The study is being carried out with the cooperation and support of the local administration and public health service. The field survey is done by two teams, each consisting of two physicians, one nurse and one technician. The subjects selected for the study have received a detailed letter giving full information about the aims of the study and its protocols, and will be invited by the survey teams to attend one of the public health centres involved in the study for up to 1.5 hours. A questionnaire is used to record respiratory symptoms and to estimate disease and risk factors for obstructive diseases of the respiratory airways. A special section of the questionnaire is devoted to the evaluation of the daily exposure to air pollutants, based on daily activity patterns, in an attempt to estimate the total human exposure.

Data will be collected by automated instrumentation and computers, so that the database will be built and updated 'on-line'. Blood determinations, for example, will be updated periodically. Monitoring of air pollutants (NO_2 , SO_2 , total suspended particulates and total hydrocarbons) will be carried out. We believe this project is distinctive in using both molecular and biochemical approaches in an epidemiological study.

PAOLO PAOLETTI
ELIO MARCHESI
FRANCO VELONA
LUIGI DONATO

CNR Institute of Clinical Physiology,
University of Pisa,
Pisa, Italy
and
ENEL Italian Electric Power Authority,
Rome, Italy

Authors by right

SIR — In proposing an explanation for the tendency of scientists to close ranks during investigations of unbecoming conduct, Louis DeFelice (*Nature* 353, 104; 1991) has provided an excellent definition of "earned authorship". Authorship is earned by activities that contribute directly and substantially to the intellectual enterprise reported in the paper. These activities would also enable an author to present orally a major part of the work to a critical audience, a familiarity with the material that might be used as a simple criterion of earned authorship in cases of doubt.

Having excluded from the "earned"

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only.

category authorship gained by other means, the question remains of how to recognize indirect contributions to a paper. The answer must be that suitable credit should be given in the acknowledgements. Some senior scientists burdened with DeFelice's "baggage of success" might well feel aggrieved at the loss of the bogus intellectual kudos of unearned authorship. So be it. In most institutes, these other activities bring their own rewards, rewards that are often denied to the junior researchers who are doing the experiments, analysing the results and writing the papers. If established scientists wish to retain a claim to earned authorship, it is within their powers to set aside sufficient time for direct participation in research. Against the loss of kudos, senior scientists could balance the retention of their scientific integrity and the avoidance of guilt by association when something goes awry.

Junior scientists would probably appreciate some ammunition with which to defend themselves against hitchhiking coauthors. Journal editors should consider adding to their instructions to authors a few lines defining earned authorship and making clear that it is the only form that is acceptable.

JOHN H. R. GEE

Department of Biological Sciences,
University College of Wales,
Aberystwyth,
Dyfed SY23 3DA, UK

Nothing new

SIR — The statement (*Nature* 354, 177; 1991) that the proposed Korean science tax is "unprecedented" is incorrect. Under Republic Act No. 5448 effective 1 January 1969, the Philippines imposed a special documentary stamp tax where

proceeds went to the National Science Development Board. The tax rate was generally 100 per cent of the then current regular documentary tax rate. A series of 15 stamps (example illustrated) ranging in face value from 2 centavos to 100 pesos was issued to prove payment of this tax (for philatelic details, see *Am. Revenuer* 31, 28-29 & 51; 1977). When (or whether) the tax and the use of the stamps ceased I do not know.

TERENCE HINES

Psychology Department,
Pace University,
Bedford Road,
Pleasantville, New York 10570, USA



Science reforms in Hungary

SIR — Christopher Anderson's article (*Nature* 352, 745; 1991) is headed "Hungarian science faces sweeping reforms". New bills for higher education and for the Hungarian Academy of Sciences that will be key factors in the development of education and science in Hungary will soon be before Parliament.

The establishment of 'sterile' research institutes in the framework of the national academies was part of Communist science policy in Eastern European countries. The artificial segregation of science and higher education on a scale unknown in Western societies still threatens Hungary with the lowest ratio of university-educated people in Europe. The academy's research institutes must become fully integrated with the universities to raise the level of higher education, because there is neither enough brain power nor money for duplication.

The all-powerful Hungarian Academy of Sciences controls scientific qualification, awards virtually all the funds for research grants, runs 52 separate 'academic' institutions, and wants to continue to do so in the future.

Many of the Hungarian universities are against the so-called academic plans including the Athenaeum Initiative, and want an independent government agency to distribute grant money coming from the taxpayers. There are also demands that the academy should relinquish its bureaucratic control over the universities through its privileged issuing of scientific degrees which are required for tenure and for appointments to chairs.

Those responsible for the future of higher education, science and research in Hungary agree with the subtitle of your report that the "universities will switch to [the] Western model", but only if the Hungarian Academy of Sciences also switches to the Western model, and its hold over the universities is abolished. Otherwise the "sweeping reforms" may well be "brushed under the carpet", and the status quo will continue to destroy what is left of science in this country after 40 years of Communist repression.

MIKLÓS KELLERMAYER

Medical University of Pécs,
7643 Pécs, Ifjúság. u. 13, Hungary

Thesis defended

SIR — However desirable, publishing papers is not the main objective of a PhD project, although it is of course intended that the research will add to knowledge and that parts of it will be publishable (L. H. Breimer and D. Mikhailidis, *Nature* 353, 789; 1991). Ex-

pecting each PhD candidate to produce at least four 'first-author' publications is completely unrealistic in a situation where projects are undertaken within a wider research programme.

Not infrequently, it is the final phase of a PhD project that is the most successful and productive period. Awarding the degree on the basis of a comprehensive thesis submitted shortly after completion of the work is surely preferable to requiring candidates to 'sit out' the long publication lag time before being allowed to submit. And what if (as sometimes happens) other investigators publish similar work as a project nears completion?

Our present style of thesis challenges the candidate to bring together the results of 3–4 years of research and to organize its presentation into a coherent account of what has been achieved. The resulting thesis is an aid to others in the laboratory and to succeeding research students in quite a different way from a set of published (and 'polished') scientific papers. The latter are unlikely to dwell on the negative results of the project, but these could be important to subsequent researchers.

The proposal by Breimer and Mikhailidis that their new system should coexist with current procedures is reminiscent of the (apocryphal) government plan to bring the United Kingdom into line with other countries and drive vehicles on the right-hand side of the road. "From January 1st, all cars will drive on the right. If, after 6 months, this experiment is successful, buses and trucks will be required to do the same."

DAVID WEITZMAN

Cardiff Institute of Higher Education,
Western Avenue, Cardiff CF5 2SG, UK

Population crisis

SIR — Fernando Orrega (*Nature* 353, 596; 1991) has been carried away by his lack of enthusiasm for the United Nations Population Fund. We certainly make mistakes — including using wrong per-capita food production figures in our 1991 *State of World Population* report. But he assumes that our mistake indicates cynicism or criminal carelessness — and that we do not accept.

We (and the Food and Agriculture Organisation, from which the data came) got the figures right in the 1990 *State of World Population*. They are not comforting. They support the idea that there is an impending food crisis in developing countries, and that rapid population growth has a lot to do with it. The World Bank — which Orrega quotes — also takes this view.

There is a similar crisis in education. Mixing up enrolment figures with children at school does not help. There are

105 million children not at school and this figure will double by the end of the century, according to UNESCO. They may or may not have been enrolled in school at one time, as Orrega claims, but they are not at school now.

The cost of providing education and health care for children is significant, even in developing countries. That is one reason why people in developing countries are having fewer children. India claims to have averted 108 million births. Does anyone — even Orrega — seriously contend that India would be better off if they had been born? That is, if India's population was now 1,000 million instead of nearly 900 million?

ALEX MARSHALL

United Nations Population Fund,
220 East 42nd Street,
New York, New York 10017, USA

Schizophrenia

SIR — Julian Leff (*Nature* 353, 693; 1991) describes how the failure of the psychiatric geneticists to confirm their own work in schizophrenia as well as the failure of others to confirm it has encouraged psychotherapists to rise again. Leff considers this movement to be mainly confined to Sweden but it also exists in the United Kingdom. Nowadays, cognitive therapists may be found actively blaming mothers for causing mental illness in their children. True biological psychiatrists are rare birds. Lip service may be paid to biological psychiatry, but underneath the blaming goes on.

Psychiatry today is, I believe, on the wrong tracks entirely. More and better neuroleptics are not the answer. Seeking the cause in each individual unmedicated case is of supreme importance. With Baruk, I would get rid of the idea of schizophrenia as a disease concept. Let us find its many causes and name them according to the metabolic fault causing the psychiatric symptoms.

GWYNETH HEMMINGS

Schizophrenia Association of Great
Britain,

Bryn Hyfryd, The Crescent,
Bangor, Gwynedd LL57 2AG, UK

Alan Blumlein

On 23 May last year (*Nature* 351, 264; 1991), we published a letter from Francis P. Thomson asking for material relevant to a proposed biography of Alan Dower Blumlein. It has since come to light that Mr Thomson is unable to produce evidence of work on such a book, that he has declined to make the material already collected available to others and that readers would be ill advised to accede to his request for biographical material. — Editor, *Nature*.

Prospects for graduates

SIR — I am a recent PhD graduate, and *Nature's* manifesto for British science (*Nature* 353, 105; 1991) does not relieve my fears about the long-term prospects in science and technology and the likelihood of permanent employment in Britain for myself and my peers. Young people are certainly turning their backs on career in science, which, as the manifesto suggests, is because the infrastructure of British science is in a mess.

But I must agree with Hopkins and Rothwell (*Nature* 353, 692; 1991) that increasing the number of research studentships will only compound the already desperate state of funding of postgraduate training, and will certainly not restore the award of a research studentship as a valuable prize, which is one of the aims of the manifesto. The underfunding of research studentships by the government-financed research councils may be a root of subsequent underfunding of research as a whole.

The devaluation of students at an early stage in their career allows a certain amount of complacency to set in at higher levels as to the real value of the researcher. Apart from low salaries, perhaps a greater fear of the PhD graduate is the fear of not having a job at all, or accepting a short-term position in a field inappropriate to one's training. A more formal study programme to complement students' research activities would help the general background training of PhD students, making them a more marketable product. This and additional theoretical training is at present very much dependent upon the enthusiasm and encouragement of a student's individual supervisor, which can be very unpredictable.

The manifesto does not go far enough on the question of job security. Why should an enthusiastic young scientist spend 6–10 years gaining two or even three specialist degrees to get only a 2–3 year temporary job offer in the end in the form of a postdoctoral fellowship? There is no other profession where so much time and energy is rewarded so poorly, with young researchers beginning their precarious careers often funded only by the good grace of charitable benefactors. There is something to commend the argument that competition breeds success, and to some extent competition by groups for grants does raise scientific standards. But to new postdoctorates, this generally sows the seeds of insecurity, disillusionment and contempt, and is no way to begin a career. The assurance of a permanent job would not lead to complacency, apathy or 'inbreeding' within groups, as postgraduates would still wish to learn new tech-

nology, to work with colleagues in different establishments and to collaborate with groups offering expertise in different disciplines.

Many young scientists will leave science in the near future not only for the easier pickings in the fields of commerce, economics and law, but also as a matter of survival. The unfortunate point is that they will have passed the most difficult hurdle of acquiring their PhD. Despite this, many people will pursue a career in science, but if science and technology in Britain are to advance into the twenty-first century, then some radical changes are needed.

PAUL EGGLETON

*Department of Biochemistry
and Cell Biology,
The Hunterian Institute,
Royal College of Surgeons of England,
London WC2A 3PN, UK*

DNA sequences

SIR — Much attention has recently been paid to the patent and intellectual property issues linked with the rapid accumulation of partial DNA sequences obtained from a large number of cDNA clones. During recent meetings, HGM 11 in London in August and Genome Sequencing in Hilton Head, South Carolina, in September, it appeared that efforts in the United States, the United Kingdom, France, Japan and elsewhere will yield several tens of thousands of cDNA sequences over the next few years. This situation had not been foreseen by the experts when the plans for the next five years of the Human Genome Initiative were formulated.

The current discussions are in my opinion mistaken, and it is time to consider the issue at the level of the ethics of knowledge. Accumulation of partial cDNA sequences is a service to the scientific community, supported by funds from foundations and government agencies, designed to facilitate and speed up acquisition of knowledge on the final product encoded by the transcripts, a limiting step in the development of diagnostic tests and therapeutic drugs. Integration of these gene landmarks together with polymorphic genetic markers within the physical map of the human genome will facilitate identification of disease genes, especially those for which no chromosomal anomaly is known.

The partial cDNA sequences represent part of our human heritage and as such should be placed under the protection of an international body such as UNESCO in order to ensure that the scientific community as a whole has

immediate unrestricted access to the data as they accumulate. International cooperation and exchanges will also be essential to limit duplication of efforts.

CHARLES AUFFRAY

*Généthon,
1 rue de l'Internationale,
91000 Evry, France and
CNRS, UPR 420 BP 8,
94801 Villejuif cedex, France*

Canada calling

SIR — The telephone and the fax machine (*Nature* 353, 286; 1991) have been both a blessing and a curse. Both have speeded up communication, the exchange of information and the development of ideas. But we receive via the telephone unsolicited nonhuman intrusions and solicitations, and the fax is a means for transmission of junk mail. But rapid transmissions by telephone and fax are here to stay, and there will probably be other changes over the coming years.

We can all thank Mr Alexander Graham Bell but, to set the record straight, the telephone was 'invented' in Canada, not the United States. Bell, a Scot, had moved to Canada because of his health. He resided in Brantford, Ontario, where he uttered his now famous words to Watson in 1876. Bell did have close ties with Boston, and many of the components for the first telephone came from there. He later moved to the United States and took out US citizenship, a prerequisite for securing a US patent.

Bell thus joins a long list of Canadians who have migrated south to achieve fame and, more importantly I suppose, their fortune. Bell did retain his Canadian ties, maintaining a summer home in Nova Scotia which is still owned by his descendants.

M. P. BRATZEL JR

*International Joint Commission,
100 Ouellette Avenue,
Windsor, Ontario N9A 6T3, Canada*

Nikola Tesla

SIR — In their 'Plea from Croatia' (*Nature* 353, 597; 1991), Ivica Vucak and Antonio Juretic inadvertently cast a light on 'cause and effect'. They cite the famous scientist Nikola Tesla (1858–1943) as a Croatian. Tesla was, indeed, proud of his native Croatian region, but he was also proud of his Serb nationality and upbringing. His father was the local Serbian Orthodox priest and his uncle a bishop. In other words, Tesla considered himself to be a Yugoslav. The final twist is that, during the Second World War, his father's parish church was razed to the ground by the Croat Fascist Ustashi.

Y. KOVACH

*38 Lebanon Park,
Twickenham, Middlesex TW1 3DG, UK*

Siberia's threatened forests

Armin Rosencranz and Antony Scott

Will Siberian forests be spared the wholesale destruction suffered by the world's tropical rainforests? The enormous changes now occurring in the former Soviet Union offer opportunities as well as challenges.

THE forests of Siberia cover 2.3 million square miles, an area the size of the continental United States¹. They represent 57 per cent of the world's coniferous forest volume and 25 per cent of the world's total inventoried wood volume². In comparison, the Amazon rainforests of Brazil are almost 50 per cent smaller³. Siberian forests help mitigate global warming, constitute a natural heritage of international importance, and drive local and national Soviet economies. Yet needless pressure is put on them by inefficient processing and harvesting techniques. Moreover, the rate of logging is likely to increase dramatically as foreign timber corporations enter into joint venture agreements with the former Soviet Union.

Just as in the Amazon rainforests, augmented development would hasten the breakdown of indigenous cultures such as the Yakut, Buriat, Khants and Mansi. By destroying their traditional sources of food and shelter and irrevocably changing their subsistence economies, increased logging would destroy the methods that these people have developed over centuries for sustainable use of boreal forests. It is essential that the allure of hard currency and modern technology does not outweigh the long-term health of the forests, the people and the local economies they support.

Foreign investment

Soviet planners are embarking on joint ventures with foreign multinational companies to help expand Siberian timber operations. Currently, 410 million m³ of wood, representing the felling of more than 4 million hectares of forest, are harvested each year in Siberia¹. Increasing this cut through joint-venture operations could help stimulate the sagging Soviet economy: investment capital is scarce, so operations can be set up quickly and efficiently only with foreign investment. Soviet planners have long exploited Siberian natural resources as a ready source of hard currency: half of the Soviet Union's 'real' money comes from export of Siberian natural resources. Export of Siberian timber now accounts for only 2.6 per cent of the Soviet Union's total foreign trade, and many Soviet interests would increase this amount. Siberian joint ventures are being negotiated with Japanese, Korean, US and European multinational timber

companies. US timber giants Louisiana Pacific, Weyerhaeuser and Georgia Pacific are known to have entered into negotiations to cut Siberian forests. Such agreements will almost certainly increase the rate of logging.

The recent completion of the Baikal Amur Mainline (BAM) railway in Siberia helps make expansion possible by alleviating some of the transport problems that had previously hindered timber extraction⁴. BAM has provided access to 1,400 million m³ of new, marketable timber⁵, and its completion permits, indeed encourages, greater exploitation of the forests. The thousands of people brought to Siberia to work on the railway are now jobless, putting pressure on officials to expedite development of other industries, such as timber. Despite the railway, most experts in international forest trade believe that large-scale joint ventures in the former Soviet Union will be difficult to broker for the next 3 or 4 years because of difficulties in transport, extensive bureaucratic red tape and the non-convertibility of the ruble³. But the biggest concern for foreign investors is political instability and the looming issue of who will own and control Siberia's forest resources.

Although prospects for large-scale investment may not be promising in the short term, some timber companies, especially from Japan and Korea, are already tapping the huge forests of the Soviet Far East. For example, the Japanese C. Itoh and Company has begun harvesting operations in Lidoga, Khabarovsk Kray, and plans to expand in 2 years. And Hyundai, Inc., of South Korea, is said to have contracted to log roughly half a million acres in the Pozharski region of Primorsky Kray.

Increased trade with the West is often held out as a solution to the Soviets' environmental (as well as economic) woes. Typical Soviet milling practices and technologies cause tremendous waste, using three times as much timber to produce a finished product as do North American and Western European companies. Soviet companies lack the technology, for example, to make use of limbs and branches of harvested trees. Moreover, many Soviet mills do not have moveable head saws, so all logs must first be milled to standard-sized planks before being remilled to the desired size. This two-step process results

in enormous waste. Western joint venture partners could alleviate much of this waste, and presumably some of the pressure to log, by introducing efficient, low-waste processing technologies.

Environmental threat

But joint ventures are more likely to exacerbate than mitigate forest damage. One would not expect that joint-venture partners would be better stewards of Siberian forests than their Soviet counterparts, given companies' destructive practices at home in the face of popular protests. Yet Soviet timber harvesting practices have already led to excessive cutting and have put a heavy burden on forest ecosystems. Ninety per cent of trees in the Soviet Union are harvested by clear or comprehensive cutting⁶, after which increased erosion causes topsoil loss and inhibits tree regeneration. Topsoil loss also leads to siltation of streams — destroying fisheries, upsetting riparian ecosystems and threatening local economies that depend on the fish. Tree regeneration is further hindered by the harsh growing conditions in Siberia: the average diameter for mature trees in Siberia is only 24 cm, and average tree growth is two to three times slower than in the rest of the Soviet Union.

Because it is the most cost-effective way of harvesting timber, clear cutting will almost certainly be the main method used by joint ventures. Indeed, the proposed Weyerhaeuser project on the coast of Khabarovsk Kray would use clear cutting to the detriment of the fragile soils there. Because much of the accessible timber in Siberia is in mountainous areas, clear cutting will lead to greatly increased soil erosion.

Unfortunately, the Soviet regulatory system will be of little help in limiting irresponsible logging practices. The gap between the Soviet Union's relatively progressive environmental statutes and their enforcement is huge⁷. The environmental consequences of projects are not formally considered and environmental impact assessments are not performed⁸. Foreign companies seem able to conduct resource extraction operations virtually unhindered. Fines for destructive practices are too small — only 100 rubles.

Goskompriroda, the State Committee on Nature Protection, may be even less effective in regulating joint ventures than it has been in controlling domestic

operations. Because it benefits from joint ventures, the government has an interest in easing the regulatory burden. For example, although one North Korean logging operation encroaches on a legally protected nature preserve, government officials overseeing the timber industry have not responded to calls by local authorities to stop the logging.

But there are practical difficulties in fining joint-venture operations as well. Because the ruble is not convertible, most joint ventures are structured on a non-cash basis. This makes collecting fines problematic. The Tyndales Production Association in Tynda, Amur Oblast, for example, is a joint venture with a North Korean timber company. North Korea gets 39 per cent of the cut timber and the Soviet Union 61 per cent. When popular protest revealed that the Koreans had over-cut tracts assigned to them, even more cutting was the result: the Koreans could pay their fines only in timber because no currency was involved in the joint-venture agreement.

Economic obstacles

The devolution of political and economic authority in the Soviet Union raises the question of who now owns — and stands to profit from — Siberia's forests. It is clear that timber profits will no longer primarily accrue to the central, or Union, budget. But it is not clear whether control will rest with the Russian Republic (of which Siberia is a part) or with more local authorities. Either outcome threatens efforts to protect the forests.

Historically, most of the revenue generated by logging has benefited the political centre of the country, rather than local Siberian economies. Not only have the bulk of the profits been skimmed off (more than 70 per cent of an operation's profit has been appropriated by the central budget) but the bulk of the economic activity generated by the timber industry has occurred outside Siberia. More than 60 per cent of the timber exported to the rest of the Soviet Union is unmilled³, as is 75 per cent of Siberian timber exported abroad⁹.

If the republics assume economic control of the forests, current policy would favour the continued export of unprocessed Siberian timber to processing plants in the European parts of the Russian Republic. The small local share in the benefits from forest harvests will fuel demand for increasing logging activity, and economic planners in distant Moscow are unlikely to examine whether their development policies undermine the sustainability of the forests. But if local interests gain control, it is not clear that the forests would fare much better. First, having so long been denied the major economic benefit of the forests,

local interests could decide to exploit the forests to the maximum extent possible. The lure of consumer goods and an increased standard of living may outweigh the need for sustainable management: there is already evidence that this is happening in some areas.

The future of Siberian forests could also be compromised by corrupt local forestry officials. Stock can be underestimated so that official allowable harvests may be similarly underestimated, allowing a black market. If control of the forests devolves to the local level but the same forestry officials remain in charge, then the profits from timber harvests could again fail to satisfy local needs and would increase calls for more cutting.

The environmental movement in Siberia has deep roots, and environmentalism all over the Soviet Union is burgeoning. But it cannot monitor joint-venture operations or demand that they adhere to certain conditions. Such operations are generally brokered in secrecy. There is no citizen review process to assess the social and environmental consequences of any project⁸, nor is *Goskompriroda* consulted. Multinationals are therefore unaccountable to local populations. Projects can be initiated with little opportunity for citizen protest because there may be no foreknowledge of the project or its impact. Environmental and citizen advocates thus have no opportunity to insist that agreements or operations accommodate environmental and social concerns.

Once operations are under way, citizens are ill-equipped to monitor operations or to compel enforcement of regulations. The difficulties of transport, the enormous scale of Siberia's forests, and the lack of pollution-monitoring equipment hinders efforts to oversee hinterland operations. Moreover, the task of organizing environmental opposition to destructive projects is thwarted by difficulties in disseminating information: photocopiers, fax machines, personal computers and even paper supplies are generally unavailable to environmental organizations. Although the Soviet environmental movement is very powerful in some respects — activists were reportedly responsible for forcing the closure of more than 1,000 environmentally degrading projects during one 6-month period — grassroots environmental groups typically lack the expertise to offer positive suggestions about development practices. An environmentalist 'success' is usually confined to shutting down or halting construction of an enterprise. The economic stagnation that results from such activism may ultimately backlash, given the urgent need for jobs and basic necessities.

It is not necessarily in the best interest of Siberian environmentalists to reject

all timber joint ventures. Foreign companies have the capital and know-how both to implement sustainable forestry practices and intensively to process harvested timber so that little is wasted. Such responsible operations would benefit the local Siberian economy and decrease pressure on the forests. They could also serve as an example to domestic timber operations. But, without knowledge of these development strategies and the economic and environmental benefits that can accrue from them, local environmentalists cannot be expected to do more than resist all development.

International implications

The international community is largely ignorant of the importance of Siberian forests. Discussions about global warming tend to focus on the need to preserve tropical rainforests, yet Siberian forests have a large role to play in mitigating global warming. It is not yet clear how large a carbon 'sink' the Siberian forests represent, but some researchers have suggested that boreal forests may account for the major part of the world's carbon sink. Siberian forests are certainly an important carbon store, so that accelerated deforestation could contribute significantly to global warming. Although research is by no means complete, forested Siberian lands may represent as much as 40,000 million tons of stored carbon¹⁰. This compares impressively with the much more publicized forests of the Amazon basin, which account for approximately 80,000 million tons of the world's stored carbon¹⁰.

The importance of Siberia's forests in the global carbon budget must be communicated to the negotiators of international agreements to limit greenhouse-gas emissions. The international community needs to recognize Siberia's forests as an economic, climatic and wilderness resource of global significance. International as well as domestic pressure must be exerted for the sustainable development of these forests. □

Armin Rosencranz and Antony Scott are at the Pacific Energy and Resources Center, 1055 Fort Cronkhite, Sausalito, California 94965, USA.

1. Barr, B. & Braden, K. *The Disappearing Russian Forest*. (Rowman and Littlefield, London, 1988).
2. Backman, C. & Waggener, T. *Summary, Working Paper 28, Center for International Trade in Forest Products*. (University of Washington, Seattle, 1990).
3. *World resources 1990-1991: A Guide to the Global Environment* (Oxford University Press, New York, 1990).
4. Cardellicchio, P., Binkley, C. & Zausaev, V. J. *Forest* **88**, 17 (1990).
5. Barr, B. *Soviet Geog.* **30**, 283-292 (1989).
6. Barr, B. H.R. *MacMillan Lectureship in Forestry* (Univ. British Columbia, Vancouver, 1989).
7. Gregory P. & Stuart, R. *Soviet Economic Structure and Performance* (Harper and Row, New York, 1986).
8. Pryde P. *Soviet Geogr.* **29**, 560 (1988).
9. Bradshaw, M. *Soviet Geogr.* **29**, 385-386 (1988).
10. Trexler, M. *Minding the World's Carbon Store* (World Resources Institute, Washington, DC, 1991).

Needed: Fetal Tissue Research

The US House of Representatives has recognised the value of studies of the transplantation of human fetal tissue. Research physicians and patients' advocates hope the US Senate will also see the light.

Experimental surgery shows that the transplantation of cells from a human fetus into the brain of an adult suffering from the tremors and rigor of Parkinson's disease can bring substantial relief to some patients whose symptoms prior to surgery included an inability to speak or move.

Promising data from Parkinson's patients, from diabetics who have received fetal islet cell transplants, and from an apparent recent success in Hurler's syndrome have renewed political lobbying by patients' groups and research organizations for a law that would overturn a four-year moratorium on federally sponsored fetal tissue transplantation studies (see *Nature*, **355**, 189; 1992).

First the Reagan and now the Bush Administration argued with unerring logic that to permit medical experimentation with fetal tissue is to encourage women to conceive and abort in order to provide suitable human material for research. It speaks well of women (and men) that even a special commission on the subject, appointed when Ronald Reagan was in the White House and George Bush was second in command, found no evidence whatsoever to support the reasoning that allegedly stands behind the moratorium.

However, actions reveal that the real reason for the moratorium was (and is) staunch opposition to abortion at any time, for any reason. This is apparent from the Reagan-Bush Administration's response to the advice it got. Despite a resounding recommendation in 1988 that the moratorium (unique in its ban on medical research for ideological reasons) be lifted, the White House simply ignored the advice of the committee which included Dr Bernadine Healy, now director of the US National Institutes of Health (NIH).

Legislation to overturn the moratorium recently passed by a majority of 274-144 in the US House of Representatives and now awaits action in the US Senate. But a presidential veto is likely to stand in the way of research even if Congress does vote in favour of the bill.

Meanwhile, much of the research in the field is taking place outside of the United States—for instance, in Sweden where 11 Parkinson's patients have been treated, in Britain where 48 transplants have been reported (though not published in detail) and in Mexico. The human experimentation that is taking place in the United States is funded completely by private money.

Although figures on the number of fetal tissue transplants varies, estimates suggest that as many as 100 patients worldwide may have undergone experimental therapy for Parkinson's and other diseases during the past three-to-five years. No one has been cured, but positive results have reinforced past promises that fetal tissue holds a special place in medicine, if only its powers can be discerned.

Conventional wisdom held that fetal tissue engrafted in the adult brain produced sufficient dopamine to bring about some relief in Parkinson's. However, recent data, reported at the annual meeting of the Society for Neuroscience last November (but not yet published) suggest another mechanism. Eugene Redmond of Yale University School of Medicine, who with independent funds has transplanted fetal tissue into 11 Parkinson's patients, found no dopamine in fetal cells at autopsy of one patient who died of the disease.

His observation, and those from animal research, lead to speculation that the fetal transplant is somehow stimulating the adult brain to resume its own dopamine production. If this proves to be the case, research rapidly will be directed toward the isolation of growth factors in fetal tissue.

In an interesting twist to current US policy, there is no ban on fetal tissue research in vitro or in animals that is comparable to the one on human fetal tissue transplantation. As a result, the NIH are currently spending \$8 million a year on research in this field.

(It is worth remembering that in the early 1970s in the United States and Europe there were intense debates about the ethics of doing any research at all on fetal tissue. The current debate on fetal tissue transplantation in humans is merely an extension of a debate about ethical issues that then, as now, was driven by deeply held views on abortion.)

To date, fetal tissue transplants in patients with diabetes, like those in Parkinson's, have been promising but not definitive. Following unsuccessful experimental surgery using cells from the pancreas of adult cadavers (graft rejection being one problem and difficulty in isolating whole adult islet cells being another), fetal tissue was a logical next choice. Subsequent research has shown that with a fetal transplant, it is not necessary to isolate insulin-producing islet cells from

other pancreatic tissue and approximately 30 diabetics in the United States are living with successfully transplanted tissue that has reduced, but not eliminated, their need for exogenous insulin.

More recently, fetal tissue transplantation was in the news when Dr Robert Nathan Slotnick of the University of California at Davis grafted cells from an aborted fetus to a fetus in utero with Hurler's syndrome, a rare genetic disorder of sugar metabolism that is invariably fatal by the age of ten. Although the operation cannot yet be proclaimed a success, the baby has been born and, so far, appears to be well.

The case of the Hurler's baby casts light on the debate about whether it is morally acceptable to use tissue from an aborted fetus to save the life of another, just as society endorses the use of tissues and organs from the deceased to save the living. The baby was born to a Baptist couple who strongly opposed abortion and still do but who, when faced with the fact that they were about to bear their third child with Hurler's syndrome, decided in favor of the transplant. (Their two other affected children have died.) Seeking guidance in the Biblical story of God's creation of Eve from Adam's rib, the baby's father has been quoted as saying, "God formed one human being from the tissue of another. Not only does God approve of this [transplantation], he himself performed the first one."

But one does not have to go back to Adam and Eve to find moral justification for fetal tissue transplantation research. The US NIH committee that called in 1988 for the end of the moratorium based its judgment on ethical analysis from philosophy and many religions. In Britain, a committee headed by Dr John Polkinghorne offered guidelines for the ethical use of tissue from aborted fetuses in medical research (see *Nature*, **340**, 327; 1989) that include provisions for informed consent and the separation of the woman's decision to abort from a decision to authorize medical use of the fetal tissue.

The ethicists who want to strike a moral balance between the rightful needs of future patients to advances in biomedicine and the need to respect human fetuses have done their job and written not only well but thoroughly on the subject. Scientists and physicians are trying to do their jobs now. The US government should not stand in the way.

Barbara J. Culliton

In the Grube in the Eocene

Leonard Krishtalka

PALAEOBIOLOGISTS attending a recent conference in Germany* were treated to a series of stunning discoveries — among them the oldest known anthropoid primates, the earliest remains of Australian placental mammals, and possibly the oldest evidence of flower pollination by an insect. The centrepiece of the meeting was, however, Grube Messel itself, a site near Darmstadt that has produced its own remarkable record of plants and animals which lived in and around a tropical Eocene lake about 49 million years ago. The Messel fossils are especially notable in two respects, making them well worthy of an on-site symposium. The organisms, including their soft parts and stomach contents, are preserved in a near-perfect state, and Gondwanan interlopers are present, including a South American anteater (edentate), an African anteater (pholidote), and a flightless, *Rhea*-like bird.

In 1859, as Charles Darwin was publishing *On the Origin of Species*, a small siderite mine was opened at Messel. Fifteen years later, the miners discovered a thick sequence of oil shales and within it the petrified remains of a gigantic crocodile, later named *Asiatosuchus*. Subsequent excavation propelled the site to fame for the abundance and extraordinary preservation of its Eocene plants, insects, bony fishes, frogs, reptiles, birds and mammals. When Allied bombers flattened Darmstadt and its Landesmuseum on the night of 12 September 1944, the collection of fossils survived, having been scattered for safekeeping. But the site almost did not survive the local government, which, in a fit of misguided utilitarianism, began to turn it into a rubbish dump in the 1980s. Two events won Messel political recognition and a temporary reprieve: an international scientific symposium in 1987, and publication in the following year of a splendid coffee-table book on the locality (*Messel — Ein Schaufenster in die Geschichte der Erde und des Lebens*; Kramer, Frankfurt, 1988). Grube Messel is now a candidate for Unesco's World Heritage list of protected sites.

Given this security, the conference organizers decided to forgo Messel myopia for a broader perspective. Three papers announced palaeontological milestones from other continents, two dating from the Eocene and one from the Cretaceous. A fossil crane-fly (*Helius botswanensis*) preserved alongside a flower on a slab of red mudstone

may be the earliest record of faithful insect pollination (D. J. Rayner, University of Witwatersrand, Johannesburg). The mudstone is from a Cretaceous crater lake in Botswana, and this marriage of insect and flower may have endured for 93 million years: the crane-fly is virtually identical to living members of the genus. The evidence for pollination is circumstantial but tempting, in that the length of the crane-fly's rostrum fits the depth of the flower.

The recovery of four anthropoid teeth from the early or middle Eocene of Algeria (M. Godinot and M. Mahboubi, Université des Sciences, Montpellier) pushes back the evolutionary divergence of higher primates by about 15 million years. Before the find, the oldest known anthropoids were from 36-million-year-old sediments in the Egyptian Fayum and elsewhere in North Africa. The four teeth represent two small species, one the size of a tarsier, the other half that size. Both are described as most similar in molar morphology to the Fayum anthropoid *Aegyptopithecus*, but I predict that the anthropoid attribution will be contentious once published.

If four putative anthropoid teeth can recalibrate primate phylogeny, a lone tooth now threatens to rewrite the historical biogeography of Australia's mammals. Fragmentary remains from a locality in Queensland place two kinds of placental mammals in Australia at least 54.6 million years ago (M. Archer, H. Godthelp and S. Hand, University of New South Wales, Sydney). One is a bat, the other an archaic herbivore (condylarth), and both pre-date the rifting of Australia from Gondwana. The critical find is the single tooth of the primitive condylarth, a flightless mammal, which gives cause to abandon the traditional picture of Australia as a continent adrift, its marsupial cargo evolving in fortunate isolation — fortunate because placental mammals failed to reach Australia before its separation in the late Eocene, about 40 million years ago. Implicit in this picture is the notion of marsupial inferiority in the face of placental competition. Now, it seems, placentals did reach an Australia still moored to Gondwana, which begs the question of what happened to the Cenozoic Australian placental mammals. The logic of the fossil record (that is, stratigraphic range data) suggests that the Australian condylarth came from South America, a continent overrun by primitive Palaeocene placental herbivores. Then again, as these discoveries reveal, the logic of the fossil record is usually the

first victim of the next field season.

Most of the speakers tackled the wherefores of the Messel fauna. Indeed, Grube Messel seems tailored for taphonomy because the site represents change in one small lake and its fauna over perhaps a million years of Eocene time. But, with only a general understanding of the site's structural geology and sedimentology, the taphonomic conclusions were merely interesting just-so stories. How did the lake form 49 million years ago? German workers (for example M. Wuttke, Referat Erdesgeschichtliche Denkmalpflege, Mainz) favour the gradual subsidence of a local basin, although an ash fall layer at Messel may imply a crater lake. All agree that much of the lake was anoxic, with the fish and crocodiles inhabiting the oxygenated surface waters. Poison gas became a taphonomic theme, being invoked to explain the death of bats, birds and insects flying over the middle of the lake.

Whatever the agent of death, the animal carcasses received a soft, undisturbed burial in the bottom muds. The birds still have their feathers, and one hedgehog (*Pholidocerus*) has a hairy mane and a scaly pad on its forehead. Preserved stomach contents reveal that one of the early, horse-like perissodactyls, *Propaleotherium*, partook of a final dessert of grapes, and that Messel's smallest bat, *Palaeochiropteryx tupaionodon*, had an exclusive diet (or at least a last supper) of moths and caddisflies (G. Richter, J. Habersetzer and G. Tröster, Forschungsinstitut Senckenberg, Frankfurt). But it is not as if there was a shortage of beetles to eat. Beetles comprise 60 per cent of all the insect fossils at Messel, and like the vertebrate remains they appear freshly minted, as if they had crawled between the sheets of shale and settled into the rock.

A big mystery is the systematics of the teleost *Thaumaturus*, one of seven fish taxa at Messel; it is known from numerous specimens and other middle-Eocene sites in Germany, France and the Ukraine. A cladistic analysis concluded that the genus was a sister group of clupeocephalans (G. Arratia, University of Kansas, Lawrence; N. Micklich, Landesmuseum, Darmstadt), a relationship other fish workers (M. Wilson, University of Alberta, Edmonton; L. Grande, Field Museum, Chicago) greeted with scepticism: "with regard to *Thaumaturus* phylogeny, the clupeocephalan fishes are a red herring".

The remaining fishes (gar, bowfin, anguillid and three percoids) do not reveal the palaeogeographic affinities of the Messel fauna (L. Grande; Chang M.-M., IVPP, Beijing; J. Gaudant, Paris). But, two of the birds — a ratite and a phorusrhachid — bespeak of a Gondwanan connection, as do the pholi-

* International Conference, Monument Grube Messel — Perspectives and Relationships, Hessisches Landesmuseum Darmstadt, 6–9 November 1991.

dotan, *Eomanis*, and the edentate, *Eurotamandua* (G. Storch, Forschungsinstitut Senckenberg, Frankfurt). When and how did they become part of the European terrestrial biota? How did marsupials get to North America in the mid-Cretaceous; South America in the Palaeocene; Messel, Antarctica, Africa and Australia in the Eocene; and Asia in the Oligocene? There were enough land bridges built and dispersal avenues plotted

in the presentations to fit everyone's preferred route and timing. For a while, it seemed, we were back in the 1960s, before vicariance biogeography and its cladistic judgements. As the worn (but still serviceable) malapropism goes, it was *déjà vu* all over again. □

Leonard Krishtalka is at The Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15213, USA.

SPECTROSCOPY

Molecular computer memory

D. Haarer

A SPECTROSCOPIC technique described in *Nature* early last year¹ for detecting individual molecules trapped in a matrix, and using them as probes of the matrix's structure, is transformed on page 335 of this issue² into a scheme for making molecular-scale memory devices. Such memory elements are sought for computers of the future. But are such computers feasible and how would they compare with, for instance, the human brain?

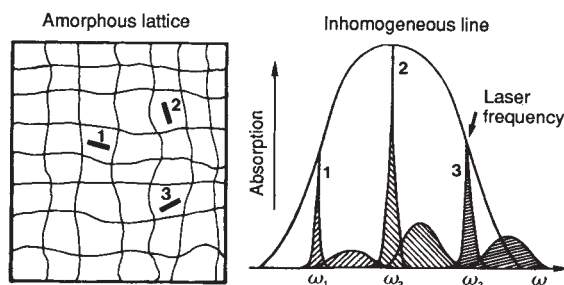
As we understand the basic concepts of biophysics better, we learn to appreciate the magnificent achievements of biological evolution, which has provided us with a brain whose capacity is of the order of 10^{15} bits. This number is based on the assumption that we have about 10^{10} neurons and about 10^4 synapses per neuron and that we can store several bits per synapse. Even though the above figure can be taken only as a very crude estimate, we know that it corresponds roughly to the content of all books ever written — not bad for a biomolecular store that operates on less than 50 watts. A computer scientist trying to configure such a gigantic storage device would have to put together a unit consuming about 100 megawatts of power.

Considerations like this have stimulated researchers in molecular science, yet, before one can think of assembling molecular systems into molecular devices, one has first to learn to deal with molecules individually. This turns out to be rather difficult, as conventional spectroscopic techniques average over a large number of molecular units, and hence provide only average values rather than 'molecular values'. But Basché and Moerner² show that much progress can be made by using very sensitive fluorescence techniques, focusing the light on a volume of roughly $10\ \mu\text{m}^3$ and detecting most photons emitted by an individual molecule by means of an efficient collection and detection system.

In a way, this study of single-molecule detection has to be compared with parallel work on trapping single ions in spe-

cial 'ion traps'³, or with that on detecting individual atoms or molecules with scanning tunnelling microscopes⁴. In all three cases the main goal is the detection of a single entity, be it ionic, atomic or molecular.

What is new about Basché and Moerner's technique, and how can it be used in science or technology? For the first



The precise absorption frequency of a molecule trapped in an amorphous matrix depends on its local environment; the cumulative effect in normal spectroscopy is to give a broad 'inhomogeneous' line. By selecting a laser frequency in the wings of the distribution, say frequency 3, one can select for a molecule in a particular site.

question, one must remember that light has a very large wavelength relative to the size of molecules and that optical experiments typically average over at least 10^7 molecules (or 10^8 atoms) if performed with a cube of side $1\ \mu\text{m}$. So optical experiments yield average values: optical spectra have broad gaussian distributions, the precise transition frequency of each molecule being determined by its environment. One common way to circumvent this problem is the use of double-resonance techniques, in which certain molecular groups are selected out of a random ensemble. The subset of molecules tuned to a particular frequency is excited by the first resonance, and then probed by a second. The basic idea of these experiments goes back to nuclear magnetic resonance experiments⁵ in which spin subgroups were addressed by local saturation or 'hole burning'.

In optical hole-burning experiments one can select molecular subensembles which are degenerate (uniform) in energy by choosing an appropriate laser

wavelength⁶. The figure shows this symbolically for an amorphous matrix doped with dye molecules. Here, different molecular environments or sites can be selected (site number three in the figure). In a typical experiment, however, 10^3 – 10^4 molecules still belong to one given subensemble.

Moerner and colleagues have shown that by making optical samples more and more dilute, one can get out of the regime of gaussian statistics: individual molecular optical transitions can be seen (especially in the wings of the statistical ensemble). These molecular absorptions appear as narrow peaks, whose width is given by the molecular relaxation times. Until recently⁷, these experiments yielded information comparable to hole-burning experiments, yet it was spectroscopy performed with single molecules rather than with subensembles.

What is new — and in crystalline systems rather unexpected — is the dynamical behaviour of the molecular spectral peaks. This behaviour is especially dramatic in amorphous polymers and occurs even at cryogenic temperatures (2 K). One can see that the frequency of the molecular transition jumps between a number of values.

These frequency jumps would be the first direct evidence of so called 'two-level systems' whose existence was put forward by P. W. Anderson⁸ and W. A. Phillips⁹ to explain some low-temperature anomalies (specific heat, sound absorption) of amorphous solids. The microscopic nature of these two-level systems is still unknown; it is believed that certain units (protons or atoms or segments of polymer chains) can tunnel between two configurations and, hence, a dye molecule in the neighbourhood will change its frequency as a consequence of these mysterious tunnelling processes in the amorphous solid. What makes the tunnelling system so interesting is the fact that the tunnelling process still occurs at the lowest possible temperatures, because it is based on a quantum-mechanical mechanism. Thus Moerner and colleagues' experiments could be the first making these tunnelling processes visible by showing well-defined frequency jumps of single molecular transitions.

Basché and Moerner now show² that the spectral change associated with those two-level systems can be induced by irradiating the absorbing molecule with laser radiation tuned to its resonant frequency; after a short time, depending on the laser power, the molecule's resonant frequency will jump. In being

able to switch and read the molecular state, the authors have the makings of a molecular memory element. They find that, in general, the molecules so altered will spontaneously revert to their original state, but they are not, in these experiments, able to control this change.

One has to discuss with more scepticism the second question of whether or not useful molecular devices can be made with such a spectroscopic scheme. In the past, molecular devices such as the 'molecular rectifier'¹⁰ have, in my opinion, been unrealistic because no viable means of writing data onto them or reading it off were devised. As long as we cannot select and switch individual molecules — with the help of some kind of Maxwell's demon — we cannot exploit the variety of molecular structures with which nature provides us. Hole-burning storage was patented long ago (1978) and the next step will be tough.

In Basché and Moerner's article the observant reader will still detect the basic limitation of the molecular scheme — the wavelength of the light which gives the spatial limit.

The only way to avoid this would be to

combine the optical experiment with a tunnelling-microscope-type device, at 2 K. So nature is still far ahead of our present spectroscopic schemes in devising molecular memories. It uses molecules much larger than perylene³ or hemi-quinones¹⁰ and has more sophisticated schemes of molecular recognition and detection. Why? Because biomolecules are tailor-made with 10⁴ molecular units and more; all in order, no imperfections. We need a little more time to achieve this. □

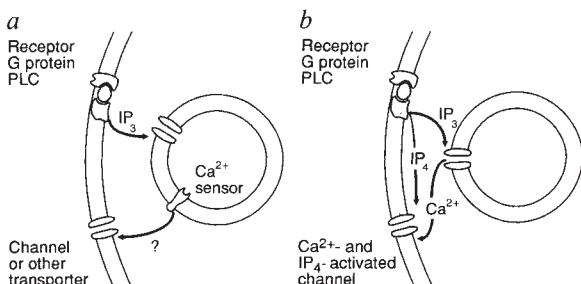
D. Haarer is in the Department of Experimental Physics, University of Bayreuth, D-8580 Bayreuth, Germany.

1. Ambrose, W. P. & Moerner, W. E. *Nature* **349**, 225–227 (1991).
2. Basché, Th. & Moerner, W. E. *Nature* **355**, 335–337 (1992).
3. Paul, W. & Steinwedel, H. Z. *Naturf.* **8A**, 448 (1953).
4. Binnig, G. et al. *Appl. Phys. Lett.* **40**, 178 (1982).
5. Bloembergen, N., Purcell, E. M. & Pound, R. V. *Phys. Rev.* **71**, 679–712 (1948).
6. Friedrich, J. & Haarer, D. *Angew. Chem. Int. Edn Eng.* **23**, 113–140 (1984).
7. Moerner, W. E. & Kador, L. *Phys. Rev. Lett.* **62**, 2535–2538 (1989).
8. Anderson, P. W. *Phil. Mag.* **25**, 1–9 (1972).
9. Phillips, W. A. J. *Low Temp. Phys.* **7**, 351–360 (1972).
10. Aviram, A. & Ratner, M. A. *Chem. Phys. Lett.* **29**, 277–283 (1974).

Controls on calcium influx

Erwin Neher

CALCIUM influx following stimulation of the phosphoinositide signal pathway is a widespread and physiologically most important phenomenon. Yet its electrophysiological basis is poorly understood — to the degree that it is hard to find a name for it. Originally dubbed ROCs, (receptor-operated calcium channels), the putative current pathways were renamed SMOCs (second messenger-operated calcium channels), a term which was itself questioned last year¹ because the role of the second messengers implicated has not been proven.



Mechanisms controlling calcium influx. *a*, Control by the filling state of intracellular calcium stores; *b*, control by Ca/IP₄. In mechanism *a*, IP₃, generated from receptor-activated phospholipase C (PLC), releases calcium from intracellular stores. A calcium sensor generates an unknown messenger when the store is depleted, the messenger inducing calcium entry through a channel or some other transporter. In mechanism *b*, the calcium released from stores, together with IP₄ (generated from IP₃ by IP₃ kinase), activates a calcium- and manganese-permeable channel.

Two reports on pages 353 and 356 of this issue^{2,3} bring new data to bear on the problem, and indicate that there is not one such mechanism, but at least two and maybe several.

Calcium influx in electrically excitable cells has been characterized in great detail. Voltage-operated calcium currents can easily be measured as their amplitudes are of the order of several hundreds of picoamperes; they need to be large because calcium has to rise significantly during the short period of an action potential. Receptor-mediated calcium influx in nonexcitable cells can by contrast last for seconds up to hours. The underlying calcium currents may be very small: in a typical cell (10 μm diameter), calcium-specific currents of only 2 pA can raise levels of intracellular Ca²⁺ at a rate of 100 nM s⁻¹. The small size makes them difficult to detect, and up to now, mainly cation channels with poor selectivity for Ca²⁺ over monovalent cations have been characterized electrophysiologically (for example, receptor-activated cation channels or channels acti-

vated by second messengers such as calcium, inositol 1,4,5-trisphosphate (Ins(1,4,5)P₃) or G proteins; see ref. 1 for review).

In some studies, inositol 1,3,4,5-tetrakisphosphate (Ins(1,3,4,5)P₄) increased Ca²⁺ influx but only in conjunction with Ins(1,4,5)P₃ (refs 4,5). From many other investigations, mainly involving calcium measurements and pharmacological manipulations, it seemed that the decisive factor for calcium entry was not the concentration of some second messenger generated by the agonist, but rather the filling state of the intracellular calcium stores⁶. In many cell types, the calcium influx phase following the Ins(1,4,5)P₃-mediated release of calcium also admits Mn²⁺ ions, which can be detected as a fluorescence quench of the Ca²⁺-indicator dye fura-2 (ref. 7).

This week's reports describe two different mechanisms of calcium influx (see figure) which may help understanding and reconciliation of the features of receptor-mediated calcium entry in nonexcitable cells. Hoth and Penner² have identified a calcium current in mast cells that is activated by depletion of intracellular calcium stores. They induced depletion of intracellular calcium stores by three independent mechanisms (EGTA, Ins(1,4,5)P₃ and ionomycin) and obtained a very characteristic calcium current in all three cases. The current has appreciable amplitude (50–100 pA) only when the free cytoplasmic concentration of calcium is buffered to very low values, similar to a current seen in lymphocytes⁸. It seems to be the most selective calcium entry pathway yet known, as neither Ba²⁺ nor Sr²⁺ can permeate. Not surprisingly, this calcium current does not seem to carry Mn²⁺.

Lückhoff and Clapham³ describe ion channels in endothelial cells that are permeable to Mn²⁺ (and presumably Ca²⁺). These channels, although not explicitly shown, are likely to have a high selectivity for divalent ions, because in the absence of divalent ions there is no inward current. Unlike the mast cell current, these channels are activated by an increase of cytoplasmic Ca²⁺. In addition, the probability of the channels in endothelial cells being open is increased by Ins(1,3,4,5)P₄, which may explain

1. Meldolesi, J. et al. *Trends pharmacol. Sci.* **12**, 289–292 (1991).
2. Hoth, M. & Penner, R. *Nature* **355**, 353–356 (1992).
3. Lückhoff, A. & Clapham, D. E. *Nature* **355**, 356–358 (1992).
4. Irvine, R. F. *FEBS Lett.* **263**, 5–9 (1990).
5. Changya, L. et al. *J. Membr. Biol.* **109**, 85–93 (1989).
6. Putney, J. W. *Cell Calcium* **11**, 611–624 (1990).
7. Sage, S. O. et al. *Biochem. J.* **258**, 923–926 (1989).
8. Lewis, R. S. & Cahalan, M. D. *Cell Reg.* **1**, 99–112 (1989).
9. Alvarez, J., Montero, M. & García-Sancho, J. *Biochem. J.* **274**, 193–197 (1991).
10. Rossier, M. F., Bird, G. St J. & Putney, J. W., Jr *Biochem. J.* **274**, 643–650 (1991).
11. Kass, G. E. N. et al. *J. biol. Chem.* **265**, 17486–17492 (1990).

earlier findings on a second-messenger role of this substance^{4,5}.

We now have a situation in which Ca^{2+} entry and Mn^{2+} influx following depletion of intracellular Ca^{2+} stores could be mediated by two different mechanisms, possibly acting in parallel in the same cell. The increase in cytoplasmic Ca^{2+} triggered by $\text{Ins}(1,4,5)\text{P}_3$ could activate the channels described by Lückhoff and Clapham (causing Ca^{2+} and Mn^{2+} influx). This would provide positive feedback for calcium entry and could be important for oscillatory changes in the concentration of intracellular Ca^{2+} . It may be that $\text{Ins}(1,3,4,5)\text{P}_4$ serves as a co-factor that determines the duration of the response; on the other hand, depletion of intracellular Ca^{2+} stores by $\text{Ins}(1,4,5)\text{P}_3$ could trigger Ca^{2+} influx through the

current described by Hoth and Penner. This Ca^{2+} entry is subject to negative feedback regulation by cytoplasmic calcium, and could be important for calcium homeostasis at some elevated level. It will be interesting to learn how the filling state of the stores might transmit a signal to the plasma membrane^{4,9,10}.

All of these mechanisms could in principle act in parallel. Indeed Mn^{2+} -permeable and Mn^{2+} -impermeable calcium-entry pathways seem to coexist in liver cells¹¹. The variety of cellular responses to external agonists may reflect which mechanism is expressed in a given cell, and to what degree. □

Erwin Neher is in the Department of Membrane Biophysics, Max-Planck-Institut für Biophysikalische Chemie, Am Fassberg, D-3400 Göttingen, Germany.

COMPTON OBSERVATORY

Gamma-ray power from 3C279

Giovanni F. Bignami

A NEW holder for the title of brightest γ -ray source in the Universe has been found by NASA's Gamma-Ray Observatory, now called the Compton Observatory¹. The source, the distant quasar 3C279, is a well known active object characterized by rapid variations in luminosity at other wavelengths.

The satellite-based Compton Observatory, named in honour of the physicist A. H. Compton who discovered the laws of electron-photon scattering, has been in operation for several months now, and is half-way through the first full γ -ray survey of the sky. Its first great success was to show that γ -ray bursters are distributed isotropically over the sky, and hence could originate outside our Galaxy (see refs 2, 3). The new discovery was made using the satellite's Energetic Gamma Ray Experiment Telescope (EGRET), a spark chamber that detects γ -rays with an energy in excess of tens of megaelectronvolts (MeV) by tracking electron-positron pairs made in γ -ray interactions. Similar in design to the earlier SAS-2 and COS-B γ -ray satellites, EGRET has vastly improved sensitivity and angular and energy resolution.

Nevertheless, the telescope was hardly pushed to get its first big result, described by Hartmann *et al.* in the latest issue of the *Astrophysical Journal Letters*¹, the chance discovery of a strong γ -ray source in the constellation of Virgo. EGRET easily pinpointed the source to within a few arcminutes, quite adequate for it to be identified with 3C279. This quasar is of the 'optically violent variable' type and is also a strong and variable X-ray source at a redshift of

about 0.5, or over 5,000 megaparsecs away (for $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$), and the first object for which so-called superluminal velocities were seen.

As such, it is much further away than the only other quasar ever seen to radiate γ -rays, 3C273, which is coincidentally in the same part of the sky. 3C273 was seen by COS-B in the late 1970s, as 3C279 would have been had its flux been as high as reported by Hartmann *et al.* — a sure indication of the quasar's variability. EGRET also measured the spectrum of 3C279: from 50 MeV to 3 GeV, it is uniformly very 'hard', well fitted by a power law with a photon index of about -2 . This indicates that the energy emitted at each frequency remains constant with increasing frequency, a tough achievement for any γ -ray source and quite unlike the spectrum of 3C273.

Curiously, a second imaging instrument aboard the Compton Observatory, COMPTEL, which is sensitive to lower-energy (MeV) γ -rays, has completely failed (at least so far) to see 3C279 (V. Schönfelder, personal communication). This may imply a significant change in the spectrum below about 10 MeV. Nevertheless, the output of 3C279 is a staggering $10^{48} \text{ erg s}^{-1}$ (or more), just in γ -rays — on a par with the total luminosity of the title holder last year for the 'brightest' object in the Universe, the infrared galaxy 10214+4724 at a redshift of 2.5 (ref. 4). The γ -ray luminosity of 3C279 is, in any case, ten times bigger than all the rest of the quasar luminosity.

Theorists are getting busy. The problem is a classic one in studies of active galactic nuclei: how can so much energy

be produced at and radiated from the active galaxy's central 'engine' (usually thought to be a massive black hole)? One must also account for the shape of the spectrum. In the case of 3C279, this problem is more formidable than usual because of its huge luminosity and hard spectrum. If the radiation is beamed along a single direction, not distributed isotropically, the problem is eased, because the total luminosity is decreased, but not drastically.

Photon-photon absorption limits the amount of energy that can escape from a bright source if the density of photons at different energies is high enough. Theorists have already had to cope with this limit in explaining the γ -rays from the weaker 3C273 source. The difficulty all depends on the size of the source region, which can be inferred from the timescale for the variability (a source can change only with a maximum rate determined by the light transit time over its dimensions). In the case of 3C279, for which rapid variability is already known at optical and X-ray wavelengths, this scale problem could be even worse than first realized, as unconfirmed reports indicate that EGRET detected significant variations over its two-week observation period. Taken at face value, this implies that those $10^{48} \text{ erg s}^{-1}$ come from a region at most a few light days across, so that it is hard to see how high-energy photons could find their way out.

The synchrotron self-Compton process is the one normally considered for the luminosity of active galactic nuclei. In this, the energy of photons is boosted in collisions with high-energy electrons according to the Compton effect (as is only appropriate for the Observatory). This process could occur in a high-energy jet streaming out of the active nucleus and so would account naturally for any beaming of the radiation and, by virtue of its instability, for the observed variability. But if there are charged particles in the jet, and if they are responsible for the γ -rays seen by EGRET, they must have unusually high energies. Their spectrum would have to be as hard as that of the cosmic-ray protons seen at the Earth. And those have been an intractable puzzle for astrophysicists for several decades. Gamma-ray quasars may be few and far between (or may have been, now that the EGRET is finally flying), but they may well represent a big step forward for both physics and astronomy. □

Giovanni F. Bignami is in the Istituto di Fisica Cosmica, Via E. Bassini 15, 20133 Milano, and University of Cassino, Italy.

- Hartmann, R. C. *et al.* *Astrophys. J.* **385**, L1-L4 (1992).
- Lindley, D. *Nature* **354**, 20-21 (1991).
- Meegan, C. A. *et al.* *Nature* **355**, 143-145 (1991).
- Rowan-Robinson, M. *et al.* *Nature* **351**, 719 (1991).

Centrosome and cell division

Eric Bailly and Michel Bornens

JUST over a century ago the centrosome was identified and dubbed the 'division organ' because of what seemed to be its main function at fertilization: organization of the first mitotic spindle of the embryo after duplication of the sperm centrosome. Although this picture has been shown to have many variations, it remains a general rule that centrosome duplication anticipates cell division. Whether this temporal sequence in any way reflects an important control on cell division has remained enigmatic¹, but the cloak of mystery has slipped a little with the appearance in *Cell* of a paper by Maniotis and Schliwa². They show that animal cells never enter mitosis following removal of the centrosome, and thereby give fresh force to the idea that centrosomes play a critical part in control of the cell cycle.

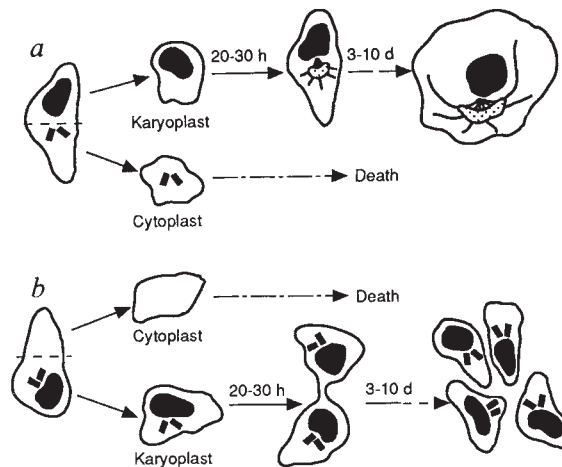
In most animal cells the centrosome typically consists of two morphologically distinct entities: a pair of cylinders (centrioles) enclosed in a complex and asymmetrical structure, the pericentriolar material from which microtubules are nucleated. The most intriguing property of this single-copy organelle is its semi-conservative form of reproduction, a daughter centriole forming and segregating with its mother. This phenomenon has been extensively investigated at the ultrastructural level but it was not satisfactorily demonstrated until recently³. In contrast, the molecular basis of centrosome reproduction has remained a matter of speculation.

To address the fundamental question of centrosome function, Maniotis and Schliwa designed microsurgical experiments to generate nucleated (karyoplast) and anucleate (cytoplast) cell fragments from cultured African green monkey BSC1 cells, with the centrosome being in either of the two pieces (see figure). Despite its unsuitability for biochemical investigations, the microsurgical approach greatly facilitates interpretation. Centrosome removal can be ascertained in each manipulation, which was not the case in previous studies⁴, enabling the authors to provide a definitive demonstration that centriole assembly cannot proceed *de novo* in mammalian somatic cells. The method also seems to be particularly well suited to unravelling the specific contribution of the centriolar components to centrosome function.

Maniotis and Schliwa found that after 20–30 hours centriole-free karyoplasts can re-establish a unique astral microtubule array focused upon a perinuclear region where the centriole location would be expected; this reinforces the

contention that centrioles are not essential for the expression of a new microtubule-organizing centre (MTOC), as implied by earlier experiments⁵. Moreover the Golgi apparatus, which was mostly eliminated along with the centrosome, regenerates and forms a compact structure at the centre of the microtubule array, testifying to the functionality of the newly expressed MTOC.

But why does this process take such a long time? The question is not easy to answer because it is not yet clear to what



Schematic representation of Maniotis and Schliwa's experiment². Nucleated fragments (karyoplasts) of post-mitotic BSC1 cells were microsurgically separated from the cytoplasmic portions (cytoplasts) that contained (a, experiment) or did not contain (b, control) the centrioles; centrioles are shown as black rectangles. The organization of both Golgi apparatus and microtubule network is represented only in the ageing centriole-free karyoplast.

extent the new MTOC is structurally and biochemically related to the former pericentriolar material. Indeed the delay between removal of the centrosome and reappearance of a radially organized microtubule network could merely reflect the time required for the biosynthesis of new MTOC components. Alternatively, this remarkable process could result from the powerful self-organization capacity of dynamic microtubules, which is thought to involve motor proteins and is apparently regulated by the cell cycle⁶. Interestingly, the resurgence of a focused microtubule array in centriole-free karyoplasts coincides in time with the first division of the control centrosome-containing karyoplasts. Yet these experiments clearly show that the centriole-free regenerated MTOC differs markedly from a classical centrosome on one fundamental point: it cannot partition and thus cannot generate the intrinsic bipolarity of the mitotic spindle.

The central issue is the identification

of the mechanism that coordinates the centrosome duplication cycle and the nuclear cycle. In essence, as for DNA, centrosomes must be accurately duplicated once and only once in each cell cycle. Either the two pathways are intrinsically dependent upon each other, some events being common to both pathways, or they are independent but coordinated through extrinsic control mechanisms which ensure a correct temporal order of events⁷.

The second idea has received the strongest experimental support. For example in early dividing embryos, where cells are not growth limited, centrosome duplication is independent of DNA synthesis and more generally of the nucleus. Moreover, the centrosome cycle can be experimentally uncoupled from the cytoplasmic clock by treating sea urchin⁸ and *Xenopus*⁹ embryos with agents that inhibit protein synthesis. A concordant result has been obtained using interphase-arrested *Xenopus* egg extracts in which procentriole budding was shown to take place¹⁰. This implies that embryos of these species possess a maternally inherited stockpile of centrosomal precursors in a ready-to-use form, a situation which does not exist in somatic cells. But the asynchrony of centrosome reproduction seen in these experiments supports the idea that normal progression through the cell cycle needs cues for proper phasing of the two cycles.

Do specific events of the centrosome cycle exert feedback control over the cell cycle? This brings us to the most exciting finding reported by Maniotis and Schliwa, the failure of centriole-free karyoplasts to initiate nuclear membrane disassembly, chromosome condensation

1. Wheatley, D. N. *The Centriole: A Central Enigma of Cell Biology* (Elsevier, Amsterdam, 1982).
2. Maniotis, A. & Schliwa, M. *Cell* **67**, 495–504 (1991).
3. Kochanski, R. S. & Borisy, G. G. *J. Cell Biol.* **110**, 1599–1605 (1990).
4. Zorn, G. A. et al. *Cell* **18**, 659–672 (1979).
5. McNiven, M. A., Wang, M. & Porter, K. R. *Cell* **37**, 753–765 (1984).
6. Verde, F., Berrez, J. M., Antony, C. & Karsenti, E. *J. Cell Biol.* **112**, 1177–1187 (1991).
7. Hartwell, L. H. & Weinert, T. A. *Science* **246**, 629–634 (1989).
8. Sluder, G., Miller, F. J., Cole, R. & Reider, C. L. *J. Cell Biol.* **110**, 2025–2032 (1990).
9. Gard, D. L., Hafezi, S., Zhang, T. & Doxsey, S. J. *J. Cell Biol.* **110**, 2033–2042 (1990).
10. Tournier, F., Cyrklaff, M., Karsenti, E. & Bornens, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9929–9933 (1991).
11. Nurse, P. *Nature* **344**, 503–508 (1990).
12. Bailly, E., Dorée, M., Nurse, P. & Bornens, M. *EMBO J.* **8**, 3985–3995 (1989).
13. Bailly, E. et al. *J. Cell Sci.* (in the press).
14. Alfa, C. E., Ducommun, B., Beach, D. & Hyams, J. S. *Nature* **347**, 680–682 (1990).

or spindle formation even though they enter and presumably complete S phase. The absence of these characteristic mitotic events points to the existence of a checkpoint mechanism that makes the activation of the M-phase controlling protein kinase p34^{cdc2} (reviewed in ref. 11) tightly dependent upon the successful duplication of the centrosome. The question then becomes how the coordination of centrosome duplication and kinase activation might be achieved. Whatever the molecular mechanisms at work, they could require direct interaction of the mitotic kinase itself with centrosomes, as outlined both by immunofluorescence studies in mammalian^{12,13}, insect (D. M. Glover, personal communication) and yeast¹⁴ cells and by cell fractionation experiments¹². Accordingly, the absence of such an interaction in centriole-free karyoplasts

would preclude the full activation of the protein kinase.

A striking property of the centrosome is that its duplication cycle encompasses all four phases of the cell cycle. In this context one might wonder why centrosome duplication starts so early in the cell cycle. A conceptually appealing hypothesis is that the discrete events of this process define a sequence of checkpoints: the failure of any one of these events would affect progression of the cell cycle by a feedback mechanism. In other words, the centrosome cycle might provide the cell with a kind of logical cog that monitors time and mass increase, thereby contributing to the coupling of cell growth and cell division. □

Eric Bailly and Michel Bornens are in the CNRS Centre de Genetique Moleculaire, 91198 Gif-sur-Yvette Cedex, France.

EARLY EARTH

Stalking the magma ocean

David J. Stevenson

PLANETS and smaller bodies are evidently differentiated, at least into a metallic core, silicate mantle and buoyant crust. Much of the differentiation occurred or began when the body was much hotter than now, back at the time of planetary formation around 4.5 thousand million years ago. It has been popular for some time to invoke a 'magma ocean' epoch when the temperatures were so high that most of the body was molten. During and immediately after this period, differentiation is thought to have been quick because of the rapidity of relative motion of silicate crystals or dense iron droplets relative to the low-viscosity magma. Recent attention has been focused on the Earth's putative magma ocean, and geologists assembled at a workshop in December* to debate evidence for and against such oceans on Earth and other bodies. This gathering showed a growing strength of opinion and argument in favour of their existence, but also a growing awareness of the lack of compelling observational evidence. It could be that the terrestrial magma ocean is merely a convenient metaphor for a complicated epoch in which the Earth's mantle can be treated neither as mostly fluid nor as mostly solid. This two-phase or multiphase problem still defies our geochemical and fluid dynamical understanding.

Historically, magma oceans were first invoked for the Moon. In the Apollo days, John Wood¹ and others developed a simple but still very successful picture

for the origin of lunar highland crust from fractional crystallization of the lunar magma ocean. At the workshop, Paul Warren (Univ. California, Los Angeles) and others discussed further tests of the lunar ocean concept, which seems to be holding up well despite persisting uncertainties about how it works geophysically. Jeff Taylor (Univ. Hawaii) discussed evidence for magma oceans on asteroids, mainly through consideration of meteorites; here, too, the evidence is good although the heat source for extensive melting remains mysterious.

These extraterrestrial examples are placid compared to the stormy violence of the proposed terrestrial magma ocean, in which convective velocities may have been of the order of a metre per second. This is the area where the workshop was most lively, and where the tension between geochemical and geophysical perspectives has previously been most evident. Rama Murthy (Univ. Minnesota) presented his provocative argument² for very high temperatures of metal-silicate fractionation during core formation, supportive of a magma ocean. His argument is that the abundances of siderophile (iron-loving) elements in the present mantle can be understood in terms of equilibrium partitioning at 3,000–3,500 °C. This met considerable opposition, most notably from Mike Drake (Univ. Arizona) and John Jones (Johnson Space Center, Houston), who argue that a more careful assessment of the chemistry of this equilibration is required. Nevertheless, Murthy has forced us to shift attention to the

possible consequences of very-high-temperature geochemistry.

Carl Agee (Harvard Univ.) presented his interesting experiments on the high-pressure melting of carbonaceous chondrites (primitive meteoritic material), and argued that the mantle we know most about (the peridotitic upper mantle) is a differentiate of the magma ocean epoch, with a higher Mg/Si ratio than the bulk Earth. The lower mantle may also be relatively iron-rich, as some have argued on different geophysical grounds, although this is a less certain prediction because of the unknown extent to which the core may have incorporated FeO-rich material from the lowermost mantle. However, Agee admits that Earth is not an assembly of carbonaceous chondrites, and more relevant experiments need to be performed.

Ted Ringwood (Australian National Univ.) and his colleagues, although not present at the workshop, have challenged the magma ocean concept by pointing out that the partitioning of trace elements (notably some rare earths) between perovskite and coexisting melt leads to a differentiation that would violate the nearly chondritic pattern of these particular elements in the upper mantle³. Geophysicists have risen to this challenge, and calculations presented by Yutaka Abe (Nagoya Univ.) and Viatcheslav Solomatov (California Inst. Technol.) and myself show that the separation of liquid from melt need not occur (as first pointed out by Tonks and Melosh⁴), at least until the crystal fraction reaches the value at which a solid matrix begins to develop.

There are hints in these calculations that differentiation does nevertheless occur in the later crystal-rich phase, that it is not of the kind feared by Ringwood and colleagues, but that it may indeed lead to a layered mantle with major element (Mg/Si, Mg/Fe) contrasts, as Agee argues. This brings us back to the longstanding, still unresolved debate on whether the mantle is compositionally layered (a debate that is often, perhaps erroneously, described as layered versus whole-mantle convection). The exciting challenge that remains is to relate our developing understanding of early Earth with our current views of mantle convection, geochemistry and core structure. □

David J. Stevenson is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.

* Workshop on the Physics and Chemistry of Magma Oceans from 1 bar to 4 Mbar, Burlingame, California, 6–8 December 1991.

1. Wood, J. et al. *Proc. 1st Lunar Science Conf.* 965–988 (1970).
2. Murthy, V. R. *Science* **253**, 303–305 (1991).
3. Ringwood, A. E. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. E.) 101–134 (Oxford University Press, 1990).
4. Tonks, W. B. & Melosh, H. J. *Origin of the Earth* (eds Newsom, H. E. & Jones, J. E.) 101–134 (Oxford University Press, 1990).

Deep heat

A LITTLE water goes a long way, if D. Deming's explanation for the formation of economically important Mississippi Valley-type lead-zinc deposits proves correct (*Geology* **20**, 83–86; 1992). Chemical alteration by hot fluids is the accepted origin of such deposits, but the source of the fluids has defied description. The coincidence of the age of deposits in the American mid-west with a period of growth of the Appalachian mountains has led some geologists to suggest that water squeezed out from under the newly elevated range could have made the 1,000-km journey to Missouri, but that still doesn't account for its temperature. Deming argues that the plate movement that built the Appalachians would also have fractured the crust to depths of 15 km or more. Fluids here could start to convect, resulting in a massive but geologically brief increase in heat flux from the Earth's interior.

Net loss

Not only are dolphins the victims of drift-net fishing, they are also caught in netting deployed to protect human bathers from sharks. Many are young calves, some with lactating females in attendance, as revealed in a survey of 250 bottlenose dolphins (*Tursiops truncatus*) caught off Natal, South Africa, between 1980 and 1988 (V. G. Cockcroft, *J. Zool., Lond.* **226**, 123–134; 1992). Devices to warn dolphins away from shark nets are ineffective, and the only sure way to reduce the toll is to remove the nets, at least during the peak period of capture. But as Cockcroft points out, that coincides with the peak period of shark capture, and a winning strategy for the dolphins could be a losing one for the tourist industry in Natal.

Six of one

COMPOUNDS in which a carbon atom is surrounded by six rather than the usual four atoms have been described. But only now, say H. Schmidbauer *et al.*, has the central carbon actually been seen (*Angew. Chem. Int. Edn Engl.* **30**, 1488–1490; 1991). The convention-defying atom is found in gold-carbon complexes; six gold atoms, each stabilized by a large aromatic ligand, form an octahedral cage around the central carbon. By using the isotopic variant ^{13}C in place of normal ^{12}C , the authors make the central atom visible to NMR. They find that the resonance is strongly overlapped by the signal from carbons in the ligands, which explains why it was not apparent when they used natural carbon (containing 1 per cent ^{13}C). Uncharacteristically moderate screening of the atom by the surrounding ions is needed to place the resonance at this frequency.

recA: from locus to lattice

Kevin McEntee

FEW proteins rival the biochemical versatility of the recA enzyme found in *Escherichia coli*. Despite its modest size (352 amino acids), it possesses an extraordinary portfolio of activities that range from homologous recombination to mutagenesis and control of gene expression during periods of cellular stress. In two papers on pages 318 and 374 of this issue^{1,2}, Steitz and colleagues provide the first detailed view of the structure of this unusual protein. By combining the structural data with genetic information, the authors offer some intriguing suggestions as to how the protein uses its structural forms to modulate its functional diversity.

Long before the biochemical properties of recA protein were known, the genetic complexities of this locus were well documented. Many mutations in the gene completely abolished recombination and conferred upon cells a profound sensitivity to killing by radiation and chemical agents that damage DNA. Moreover, the mutants were unable to mount a cellular stress response, the SOS response, following exposure to these same agents. Although most mutants showed these wide-ranging defects, another class of mutation left cells partially disabled by blocking only a subset of these functions. A third class, represented by the *Tif-1* allele, caused cells constitutively to express the SOS response in the absence of any cellular insult³.

How could a single protein be involved in such disparate cell functions? Biochemical analysis⁴ suggested a solution to the puzzle, while, at the same time, raising several new questions. *In vitro*, purified recA protein promotes the swapping of DNA strands between homologous partners. The best-studied model system involves three strands: a circular single-strand and a linear duplex. This reaction has been separated into several distinct steps which involve binding of recA protein to the single-stranded component to produce a nucleoprotein complex followed by binding of a duplex DNA molecule to the complex. Presumably, this second binding event is initially stabilized exclusively by protein-DNA interactions, with the duplex being held close to the location of the single-stranded component in the filament. The DNA molecules are brought into sequence register by a mechanism that is poorly understood but which involves extensive unwinding of the duplex recipient and possible formation of a triple-stranded DNA helix. By using DNAs substituted with 7-

deazaguanine, Jain *et al.*⁵ have cleverly shown that the structure of this intermediate differs from molecular models of triplex DNA in not requiring base pairing through the N7 position of this purine. Once homologous alignment has been achieved, recA protein must resolve the topological problems associated with unwinding the old partners and interwinding the new.

This amount of enzymatic dexterity would satisfy most proteins, but recA possesses an obviously distinct, yet equally remarkable activity that accounts for its gene regulatory function. Eschewing the vast majority of cellular proteins, it recognizes a select group of phage and cellular repressors (and at least one protein needed for damage-induced mutagenesis) and promotes their autodigestion⁶. Like its strand-exchange function, this coprotease activity resides in a recA filament and requires ATP. Although recA protein hydrolyses ATP in the presence of single-stranded DNA, hydrolysis is not mandatory for its coprotease activity. Substantial strand exchange proceeds in the presence of the poorly cleaved analogue, ATP- γ -S (ref. 7), but hydrolysis clearly stimulates the extent of this reaction and may be necessary for driving strand exchange through regions of sequence heterology⁸.

There is considerable evidence that ATP binding results in a conformational change in the polypeptide, enhancing its DNA binding. Likewise, DNA binding seems to stabilize a form of recA protein with a higher binding affinity for nucleotides. These structural changes appear not to involve large scale rearrangements of the protein, but are confined to smaller domains or regions.

At first glance, the 2.3-Å crystalline structure of recA protein¹ resembles the recA DNA filaments mentioned earlier. Monomers of recA form an extended helical polymer that is stabilized in large part by extensive contacts on two separate surfaces of each subunit. Each monomer consists of a large central domain, comprising nearly two-thirds of the mass of the protein, flanked by distinct N- and C-terminal subdomains. Not surprisingly, this central core contains the critical residues for nucleotide binding and hydrolysis as well as two disordered loops located near the polymer axis that are implicated in DNA binding. By diffusing ADP into crystals of recA protein, Story and Steitz² are able to position the nucleotide within the large central domain. Not unexpectedly, the α - and β -phosphates bind a loop region containing the highly conserved

Walker motif (G/XXXXGKT/S) found in a wide variety of nucleotide-binding proteins. By model building the γ -phosphate into this structure, it is possible to identify other residues that are important for protein-nucleotide interaction: aspartate-144 and glutamate-96 are proposed to serve directly in phosphodiester bond cleavage. By analogy with the structural elements of the GTP binding domain of the *ras* p21 oncoprotein, the authors propose that glutamine-194 may play a part in coupling ATP and DNA binding in a high-affinity binding form of the *recA* protein. Consistent with the proposed essential role of each of these residues, each shows a remarkably high degree of conservation among more than two dozen *recA* protein sequences.

The crystal structure provides some intriguing insights into the mechanism of SOS regulation by *recA* protein. Residues which by genetic studies are known to effect cleavage of temperate phage repressors (λ and $\phi 80$), as well as cellular LexA and UmuD proteins, are localized to a 'notch' formed on the outer surface of the polymer between adjacent monomers. The model is simple: repressors or UmuD protein bind to the *recA* polymer within this intersubunit region and are stabilized by interaction with each of the neighbouring monomers. It seems reasonable to assume that distinct residues within this binding domain interact with different target proteins, thereby accounting for the selective loss of induction seen with some *recA* mutations.

Within crystals of *recA* protein, filaments are closely packed in parallel arrays stabilized by specific contacts between N-terminal residues of monomers in one filament and C-terminal residues of subunits in neighbouring filaments. Once again, by examining the genetic data, Story *et al.*¹ have made the interesting observation that residues known to be affected by the *Tif* mutation and other protease constitutive alleles are close to or within these contact regions. Thus they propose that the *Tif*-1 mutation may shift the equilibrium away from multipolymeric structures or bundles and favour the free or dissociated filament solution. Because bundles appear to be competing structural forms of *recA* protein that do not participate in DNA binding, the *Tif* mutation could be acting by increasing the effective concentrations of *recA* protein available for substrate binding. *In vivo*, the consequence could be the constitutive activation of *recA* protein through its binding to 'unnatural' substrates such as small polynucleotides or replication gaps. Although such a mechanism remains highly speculative, both *in vitro* and *in vivo*, this structural mechanism for mod-

ulating enzyme function may be applicable to other protein systems with diverse oligomeric forms. Such a mechanism also explains some of the reported biochemical and kinetic properties of *recA* (*Tif*) protein. Moreover, mutations that increase the stability of bundles might be expected to manifest a *recA*⁻ phenotype *in vivo*. Such may be the mechanism of the *recA629* mutation which was selected as an intragenic suppressor of *Tif*⁹.

With this said, the question can be asked how similar this X-ray structure of *recA* protein is to the active *recA* nucleoprotein complex in solution. One of the difficulties in answering it is that it has not been possible to model precisely the position of DNA in the structure. It seems likely that activated *recA* would preserve the extensive subunit contacts found in the crystal structure, but contacts between ATP and the protein could differ between the two forms of the enzyme. Although the crystal structure implies that tyrosine-264 makes no specific contact with nucleotide, this conclusion is not compatible with photoaffinity-

ATOM OPTICS

Atoms brought to a new focus

Christopher Foot

ATOMS are usually thought of as particles — like little hard spheres — but under certain circumstances they behave like waves and undergo diffraction, interference and other phenomena familiar in experiments with light. This wave-particle duality has been discussed since the earliest days of quantum mechanics and is well known for photons and electrons. Recently there has been great interest in developing analogues of optical elements that rely on the wave prop-

erties of neutral atoms, and O. Carnal *et al.* (*Phys. Rev. Lett.* **67**, 3231–3234; 1991) have now succeeded in imaging and focusing atoms using a 'Fresnel zone plate'.

The Fresnel zone plate used by Carnal *et al.* (Fig. 1) comprises concentric opaque rings made by microfabrication from a thin gold foil; the radius of the outer ring is only 100 μm . Atoms are diffracted as they pass through the gaps between the rings so that at any point after the zone plate there are matter waves which can interfere, either constructively or destructively. A zone plate is devised to block the path of all diffracted waves that would destructively interfere at the chosen image point I (Fig. 2). The inner and outer radii of the rings are chosen such that all the allowed paths give a positive contribution to the sum of amplitudes at I and maximize the intensity. It does not actually matter much whether one chooses to block the even or the odd zones, especially in the new experiment in which the centre of the zone plate was blocked. (The details of Fresnel's theory of diffraction are given in many textbooks on optics, for example E. Hecht's

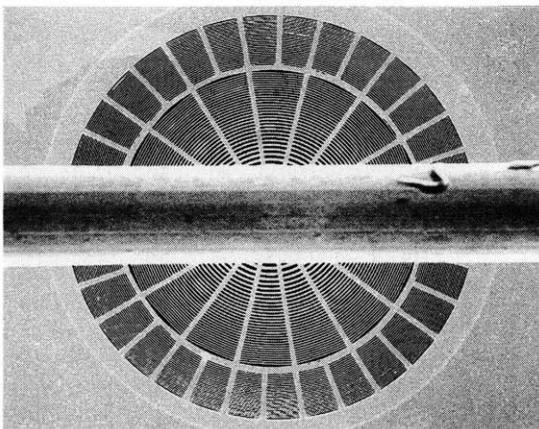


FIG. 1 Scanning electron micrograph of the Fresnel zone plate, just 200 μm across, used by Carnal *et al.* to focus a beam of helium atoms. The 50- μm -thick wire across the middle blocks the central opening of the lens to exclude the zeroth-order diffraction peak and so improve the contrast in the image centre.

1. Story, R. M., Weber, I. T. & Steitz, T. A. *Nature* **355**, 318–325 (1992).
2. Story, R. M. & Steitz, T. A. *Nature* **355**, 374–376 (1992).
3. Witkin, E. M. *Bacteriol. Rev.* **40**, 869–907 (1976).
4. Roca, A. I. & Cox, M. M. *Crit. Rev. Biochem. molec. Biol.* **25**, 415–456 (1990).
5. Jain, S. K. *et al.* *J. biol. Chem.* (in the press).
6. Little, J. W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1375–1379 (1984).
7. Menetski, J. P., Bear, D. G. & Kowalczykowski, S. C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 21–25 (1990).
8. Roselli, W. & Stasiak, A. *EMBO J.* **10**, 4391–4396 (1991).
9. Knight, K. L. *et al.* *J. biol. Chem.* **269**, 11279–11283 (1984).
10. Freitag, N. E. & McEntee, K. *J. biol. Chem.* **266**, 7058–7066 (1991).

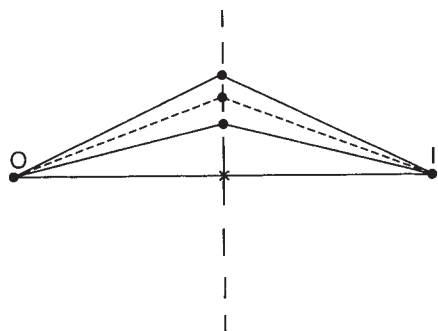


FIG. 2 Geometry of a zone plate. A cylindrical beam can be considered to pass through a series of concentric circular zones (Fresnel zone). By blocking off the odd-numbered zones (or even-numbered ones) with opaque rings, one can increase the amplitude at the image plane I produced by waves from an object at O by allowing only paths that interfere constructively.

Optics, 2nd edition, published by Addison-Wesley in 1989).

Curiously, the total amplitude and the intensity, proportional to the square of the amplitude, are much larger with the zone plate than in its absence. Thus the presence of the opaque rings actually increases the intensity at the image point over what would be observed if there were no zone plate — a rather counter-intuitive result. This prediction was verified by Lord Rayleigh in 1871 by making a zone plate for visible light (the first zone radius is of order a millimetre). Zone plates are not generally used for wavelengths for which suitable materials exist for making conventional lenses, but Fresnel lenses which are based on an extension of the same principle are commonly used in overhead projectors because they provide economical large-area lenses for collecting light from the bulb.

Carnal *et al.* used a commercial Fresnel lens suitable for focusing X-rays in their new experiment with atom beams. A beam of helium atoms, excited into a long-lived (metastable) excited state for ease of detection, was directed through an object structure — a single slit or pair of slits — and then diffracted by the zone plate to form an image (Fig. 3). The distribution of the focused atoms in the image plane was recorded by a detector sensitive to metastable atoms.

The authors found the image to be 7/8 of the object size, in proportion to the ratio of the image-to-object distances just as expected for any lens. To obtain clear images, they needed to make several refinements. The atoms all had to have similar de Broglie wavelengths (to make the beam 'monochromatic').

Carnal *et al.* achieved this by using a supersonic source of helium which produces atoms with only a small spread in velocities, and by exciting these atoms to the metastable state with a collinear electron beam. Moreover, the temperature of the source could be varied from room temperature down to 12 K by cooling with liquid helium.

This changes the mean atomic velocity, which is inversely proportional to the wavelength, so that the authors verified that a well-defined image is formed only at a particular wavelength for a particular distance of the detector from the lens. A further refinement was to place a wire across the centre of the zone plate, greatly increasing the contrast by blocking the undiffracted zeroth order of the beam which would give a high background intensity. Nonetheless, 6 per cent of the metastable atom flux before the zone plate was focused into the image. The radial bars in the zone plate, which hold the rings in place, do not influence the imaging properties.

This first experiment has shown that a Fresnel zone plate is a useful element in atomic optics. It has a great advantage over other methods for focusing atoms, based on static magnetic or electric fields, as it depends only on the de Broglie wavelength of the atoms, not on

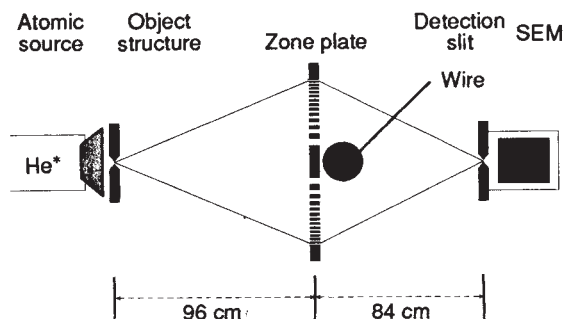


FIG. 3 The experiment of Carnal *et al.* The metastable helium atoms (He^*) pass through an object structure before being focused by a Fresnel zone plate, and cause the moveable detector (a secondary electron multiplier; SEM) to give out a small pulse of charge on each impact.

their internal properties, and so is not specific to any particular atom.

The performance of the system could be improved by using a more monoenergetic atom beam produced by laser cooling of the atoms, and also the intensity could be increased by pushing current microfabrication techniques to their limit to make a larger zone plate. Carnal and colleagues suggest that finely focused atomic beams could achieve a spatial resolution of a few nanometres and that matter waves produced by this technique could be used to probe material surfaces. □

Christopher Foot is in the Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

Economy of scale

ACCORDING to one theory of smell, our noses respond to the various frequencies at which smelly molecules vibrate. Last week, Daedalus proposed that micro-organisms work the same way. They sense their chemical environment by reacting to the vibrational modes of nearby molecules. If so, says Daedalus, they should also react to radiation of the same frequency or set of frequencies as those molecular vibrations.

Unfortunately, this radiation would be in the far infrared or higher microwave region, which is strongly absorbed by the water in which the organisms live. So Daedalus proposes to irradiate them with visible light, modulated at the relevant far infrared frequencies. This is easily generated by modern optical methods and travels readily through air and water. Its interaction with the organisms' receptors, however, will be strongly non-linear. The light will be demodulated, the modulation frequency will be reinstated on the surface of the organisms, and will trigger their chemoreceptors.

At a stroke, our control over the little beasts will be wonderfully enhanced. Genetic engineers trying to persuade a group of reluctant *E. coli* to swallow some crafty plasmid will merely have to irradiate the colony with light carrying the correct trigger frequencies, when they will all obediently gulp it down. Doctors faced with a staphylococcal infection resistant to the usual antibiotics could shine light on it designed to trigger the cell wall opening response of the creatures, when they would leak to death. Protozoa such as paramecia, which navigate on chemical clues, could be steered around at will. 'Tickled' with light designed to imitate a noxious chemical, they would immediately turn away; a second flash mimicking a nutrient particle would turn them towards it. Under light simulating the attachment signal of a second organism, they would bind to an object presented to them. In fact they could be taken over completely.

Nanotechnologists will welcome these willing workers. Under properly programmed light, a paramecium could seize a fine copper wire, and carry out point-to-point wiring on the most complex microchip. A whole squad of them in the same programmed light could even wire up a big matrix of chips simultaneously, each conducting an identical 'walk' around its local environment. Other optical command sequences could instruct the creatures in microweaving, micro-engraving with tiny diamond scalpels, and even the assembly, operation and improvement of micromachines. The industrial revolution could be repeated on a microscopic scale.

David Jones

AIDS and malaria experiments

SIR — Charles Gilks (*Nature* 354, 262; 1991) suggests that AIDS may have initially entered the human population by direct inoculation into human prisoner volunteers with malaria infected blood which may have been contaminated with the primate precursors of the human AIDS virus. Gilks states "... my theory is testable. The records of human volunteer experiments, particularly those involving prisoners in the United States, can be checked to see if any further, unpublished studies took place." But Gilks' theory is *not* testable, and will ultimately be added to the increasing list of unproven theories proposed to explain the origin of the AIDS epidemic.

As a clinical associate at the National Institutes of Health (NIH) in 1962, I and my colleagues performed monkey and human malaria studies in prisoner volunteers at the Atlanta penitentiary, discovering the indirect fluorescent antibody test for malaria (for example, S. F. K. *et al. Science* 135, 1130; 1962). All research programmes at the penitentiary had to receive prior approval from the NIH ethical behaviour board, which never allowed direct primate blood inoculation into humans. Our experiments always and only involved mosquito transmission of monkey and human malaria into volunteers. It is a matter of scientific fact that AIDS cannot be transmitted by mosquitoes or other arthropods.

Gilks' suggestion that "material from several of the original experiments could still exist and could be tested for the presence of retroviruses" is unfortunately untrue. In the case of the prisoner volunteer studies carried out at the Atlanta penitentiary up to 1962 serum samples from the volunteers, to the best of my inquiries, no longer exist after over a 30-year lapse. Blood from monkeys infected with *Plasmodium gonderi* used in prisoner volunteer studies in 1961 and 1966 by other investigators still does exist at the Centers for Disease Control. But, as in our studies, these infected *P. gonderi* monkey blood samples were used only in failed attempted mosquito transmission studies, never by direct human inoculation, which has always been prohibited in the United States.

SANFORD F. KUVIN

44 Coconut Row,
Palm Beach, Florida 33480, USA

SIR — Gilks¹ has offered a provocative explanation for the origin of the human immunodeficiency viruses in which researchers may have encouraged animal retroviruses to jump the species gulf by injecting simian blood into humans, including themselves. If a long and possibly fruitless search for surviving evi-

dence is warranted on the strength of his arguments, we must also consider alternative explanations for the origin of HIV involving clinical research.

After the First World War moderate professional opinion was deeply divided over the question of whether the physical and mental disabilities of mid- and late life could be postponed or averted by boosting sex-hormone production. Among the treatments offered for rejuvenation, testicular grafts were regarded as particularly valuable and effective, hundreds having been performed in Europe and the United States by the mid-1920s²⁻⁴. There were always more willing patients than donor organs from healthy young men and, as knowledge of transplant biology was primeval, surgeons frequently used animal testicles, preferably from apes and monkeys.

Common chimpanzees were favoured donors, not only because of a genetic affinity with man but also on account of their impressively large testicles. Few details have survived about the history and condition of these animals or the much larger numbers of monkeys, although many had probably originated in the French colonies of West Africa. One testicle was removed, sliced and grafted either to the abdominal rectus muscle or to join the organs resident in the patient's scrotal pouch. In either case, the grafts effectively inoculated him with whatever infectious agents they were carrying. When synthetic testosterone became available in 1935 and careful testing showed that rejuvenation therapy had been wishful thinking, testicular grafting became disreputable and former patients threatened their doctors with suits. Nevertheless, administration of animal sex organs by grafting or injection did not completely die out.

Although we may never know precisely how many people received simian testicular grafts they far outnumber those treated with malarial blood. But by drawing attention to the possibility of a biomedical origin of HIV, however speculatively, we risk the possibility of creating new scapegoats. Many of the rejuvenators were well-intentioned, if misguided, doctors and it would be regrettable if they or the malaria researchers were to be charged with the grave responsibility for the AIDS epidemic on the basis of circumstantial evidence.

ROGER G. GOSDEN

Department of Physiology,
University of Edinburgh,
Edinburgh EH8 9AG, UK

1. Gilks, C. *Nature* 354, 262 (1991).
2. Stanley, L. L. *Endocrinology* 6, 787-794 (1922).
3. Hoskins, R. G. *Endocrinology* 9, 277-296 (1925).
4. Hamilton, D. W. *The Monkey Gland Affair* (Chatto and Windus, London, 1986).

Neutral terms

SIR — Woese *et al.*¹ claim that archaeobacteria's most recent common ancestor appears to have been with eukaryotes, not with eubacteria. This stems from rooting the ribosomal RNA tree in the eubacterial branch by comparisons of two pairs of paralogous protein genes. The reason is that each archaeobacterial protein in the reported studies has greater similarity to its eukaryotic than to its eubacterial homologues. But several examples of the opposite exist. Furthermore, a eubacterial ATPase isolated from the eubacterium *Thermus thermophilus*² more resembles the archaeobacterial and eukaryotic ATPases than the *Escherichia coli* ATPase cited by Woese *et al.* More strikingly, an archaeobacterial glutamate dehydrogenase specifically related to its homologue from 'higher' eukaryotes, whereas its eubacterial homologue is related to 'lower' eukaryotes³. This indicates that protein phylogenetic trees cannot be used at the moment to root with confidence the ribosomal RNA tree.

Two possibilities remain: either the root is in the eubacterial or archaeobacterial branch and the last common ancestor was prokaryotic-like, or the root is in the eukaryotic branch and it resembled more proto-eukaryotes. The first hypothesis fits with the theory that life originated at high temperature, as the common ancestor of all prokaryotes was probably a thermophile¹, but the second fits better with the antiquity of introns and the 'RNA world' hypothesis. I suggest that prokaryotes arose as a specific adaptation to thermophily by reductive evolution from a common mesophilic ancestor to pro- and eukaryotes.

Both the traditional nomenclature⁴ and that proposed by Woese *et al.* are biased in favour of the first hypothesis. The terms 'pro' and 'eu' suppose that prokaryotes are primitive compared to eukaryotes, and the term 'archaea' suggests that archaeobacteria are the more ancient life forms. We should be looking for nomenclatures which avoid answering unsolved questions. Otherwise, if prokaryotes turn out to have originated from primitive eukaryotes by reductive evolution, we will have to call them post-karyotes! I therefore propose the neutral terms akaryotes (without nucleus) and synkaryotes (with nucleus) for a natural classification without prejudice.

PATRICK FORTERRÉ

Institut de Genetique et Microbiologie,
Université Paris-Sud, Bâtiment 409,
91405 Orsay Cedex, France

1. Woese, C. R. *et al. Proc. natn. Acad. Sci. U.S.A.* 87, 4576 (1990).
2. Yokohama, K. *J. biol. Chem.* 265, 21940-21950 (1990).
3. Benachenhou, N. & Baldacci, G. *Molec. gen. Genet.* (in the press).
4. Mayr, E. *Nature* 348, 491 (1990).

Climate and landscape response

SIR — That an increase in regional elevation could have promoted global cooling during the Late Cenozoic is one of several proposed mechanisms that may have contributed to the initiation of extensive Pleistocene glaciation. Molnar and England have pointed out¹ that the validity of this model, which has regained some favour², depends on the reality of Late Cenozoic surface uplift. They also argue that the Late Cenozoic climate changes which such apparent uplift is thought to explain, could themselves “cause the same changes in the landscape and fossil record as regional uplift would”. Although we acknowledge that it is important to be aware of circular arguments in attempting to establish lines of cause and effect in complex environmental systems, Molnar and England fail to justify their assertion that Late Cenozoic climate change and surface uplift could elicit an equivalent geomorphological response at the regional scale.

Can increases in denudation rates brought about by climate change cause rapid incision of high plateaus and the formation of precipitous mountain relief? Molnar and England assume that the onset of glaciation necessarily leads to a significant increase in denudation rates, yet they provide no evidence in support of this crucial component of their hypothesis. None of the three reports on the relationship between climate and sediment yield that they cite considers glacial regimes in detail (and none allows adequately for the effects of local relief). The limited data that are available in fact indicate that denudation rates in glaciated catchments are broadly comparable to those in non-glaciated basins with similar local relief and precipitation. Indeed, maximum erosion rates appear to be associated with maximum precipitation rather than the proportion of a basin that is glaciated³.

We are unconvinced as to the impact of the increase in ‘storminess’ on denudation rates conjectured in ref. 1, even if such a climate change did occur in the Late Cenozoic. Although there is

abundant evidence that the magnitude-frequency characteristics of runoff can significantly affect rates of sediment transport, at least where sediment availability is adequate, studies of present sediment and solute yield also demonstrate that local relief exerts a critical control over denudation rates⁴. An increase in local relief, and consequently stream-channel and hill-slope gradients, in response to tectonic uplift should have a much greater impact on denudation rates than a change to a stormier climate, the more so because the associated increase in elevation will also promote more orographic precipitation.

M. A. SUMMERFIELD

*Department of Geography,
University of Edinburgh,
Edinburgh EH8 9XP, UK*

M. P. KIRKBRIDE

*Department of Geography,
University of Dundee,
Dundee DD1 4HN, UK*

MOLNAR AND ENGLAND REPLY — We suggested¹ that the various processes that created the evidence commonly used to infer recent uplift of mountain ranges could have been induced by climate change. Summerfield and Kirkbride take issue with the suggestion that climate change could affect the processes controlling rates and styles of erosion in the same way and to the degree that uplift presumably can. They can make two points: the first we consider to be minor, but the other reveals a very different perspective on how to examine what is important.

Their first comment, about glacial erosion, exaggerates what we wrote. Like others^{5–8}, we do suspect that alpine glacial erosion is rapid, and therefore that in the past few million years it may have been more rapid at high elevations than other forms of erosion were earlier. Uncertainties as to whether or not glacial erosion has been rapid, however, do not negate the suggestion that climate change has played a major role in creating the many landforms, glacial and non-glacial, commonly used to infer recent uplift of mountain ranges.

Underlying their second objection, concerning denudation rates, is the widely held belief that ‘juvenile topography’ implies recent uplift. It is reasonable to assume that abrupt uplift of terrains above the base levels of their drainage systems should cause increases in erosion rates, denudation rates and relief. The commonly made, converse step of inferring recent uplift from the existence of high relief, however, depends on the assumption that only uplift of the terrain can cause the erosive and denudational processes to increase. Perhaps this

assumption is correct, but to demonstrate its validity requires a quantitative understanding of the processes that cause erosion and sediment transport and of the factors that affect these processes, not simply an observed correlation followed by an assertion about cause and effect.

To buttress the claim that rapid denudation is the result to uplift, Summerfield and Kirkbride cite the well known correlation⁴ of relief with high sediment yield, which when scaled by the area of the drainage basins is called the ‘denudation rate’. Again, we question Summerfield and Kirkbride’s assignments of cause to uplift and of effect to rapid denudation drawn from this correlation. First, the use of sediment transport as a surrogate for denudation rates has been discouraged repeatedly, except in those studies (such as ref. 9) designed specifically to understand the processes that govern both denudation and sediment transport. Meade¹⁰ and Trimble¹¹ have shown clearly how the alteration of the landscape by humans has made sediment transport a very misleading measure of denudation. Church and Ryder¹² deduced that much of the sediment transported by rivers through glacially excavated regions was actually eroded by the glaciers and was subsequently stored in the basins for thousands of years. Thus, such ‘denudation rates’ may give very misleading measures of both current and Late Cenozoic rates of denudation.

Second, what is the evidence or logic that makes the assertion that “local relief exerts a critical control over denudation rates” more defensible than, for instance, that high local relief is a consequence of rapid sediment transport? More importantly, the commonly made assignment of the cause of rapid sediment transport to a state, high relief, shifts the focus of scientific investigation away from the study of the processes by which erosion occurs and sediment is transported. Without understanding these processes, we cannot predict quantitatively how either climate change or uplift might cause an increase in erosion rates or relief.

We certainly do not presume that tectonic processes have played no role in creating relief, but we insist that what is known about geomorphological processes permits climate change to be the cause of the ‘juvenile topography’ and rapid sediment transport that have commonly been attributed to recent uplift.

PETER MOLNAR

*Department of Earth, Atmospheric and
Planetary Sciences,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA*

PHILIP ENGLAND

*Department of Earth Sciences,
University of Oxford, Oxford OX1 3PR, UK*

1. Molnar, P. & England, P. *Nature* **346**, 29–34 (1990).
2. Ruddiman, W. F. & Raymo, M. E. *Phil. Trans. R. Soc.* **B318**, 411–430 (1988).
3. Hicks, D. M., McSaveney, M. J. & Chinn, T. J. *H. Arct. Alp. res.* **22**, 26–42 (1990).
4. Ahnert, F. *Am. J. Sci.* **268**, 243–263 (1970).
5. Embleton, C. & King, C. A. M. *Glacial Geomorphology* (Edward Arnold, London, 1975).
6. Drewry, D. *Glacial Geologic Processes* (Edward Arnold, London, 1986).
7. Harbor, J. M., Hallet, B. & Raymond, C. F. *Nature* **333**, 347–349 (1988).
8. Wahrhaftig, C. & Cox, A. *Geol. Soc. Am. Bull.* **70**, 383–436 (1959).
9. Church, M. & Slaymaker, O. *Nature* **337**, 452 (1989).
10. Meade, R. H. *Geol. Soc. Am. Bull.* **80**, 1265 (1969).
11. Trimble, S. W. *Science* **188**, 1207–1208 (1975); *Am. J. Sci.* **265**, 876–887 (1977).
12. Church, M. & Ryder, J. M. *Geol. Soc. Am. Bull.* **83**, 3072–3095 (1972).

Home Sweet Home

R. V. Jones

Echoes of War: The Story of H₂S Radar. By Bernard Lovell. *Adam Hilger/AIP: 1991.* Pp. 287. £17.50, \$30.

THE development of airborne radar in Britain between 1937 and 1945 was a triumph of science and technology. It was achieved by a small band of scientists and engineers working under the spur of war with officers of the Royal Air Force. The pre-war effort was led by E. G. Bowen, who described it in *Radar Days* (Hilger, 1987).

Bowen's team, which included Robert Hanbury Brown, worked mainly on 1.5-metres wavelength, which was the shortest on which substantial radio power could then be generated; and the team's successes culminated in two pieces of equipment, both remarkable for their time — AI Mark IV for detecting bomber from fighter aircraft, and ASV Mark II for detecting surfaced U-boats from the air. It had always been clear, though, that much higher resolution could be achieved at centimetre wavelengths, if only these could be generated with sufficient power. This problem was posed to the scientific recruits who flooded in after September 1939, and it was brilliantly solved by the invention of the cavity magnetron by J. T. Randall and H. A. H. Boot in Birmingham.

The first application of the cavity magnetron was to night-fighting, and in the course of trials it was noticed that echoes were received not only from aircraft but also from towns, whose juxtaposed buildings sometimes provided strong echoes by corner-reflection. The phenomenon had been noted by Bowen's team in 1938, but their suggestion that this could provide a means of detecting towns was not followed up, perhaps because of the reluctance to contemplate the bombing of civilian populations.

By 1941, though, attitudes had changed. Britain itself had been heavily bombed, and reluctant scientists were now ready to do anything to shorten the war and minimize the number of casualties. In addition, Bomber Command, hitherto complacent about its ability to find targets without scientific aids, now realized that its navigational techniques were very inadequate. The task of developing airborne radar as a bombing aid was therefore given to a team led by P. I. Dee and responsibility for the airborne equipment was given to Bernard Lovell, who now describes the work, and its outcomes over both Germany and the Atlantic, in *Echoes of War*.

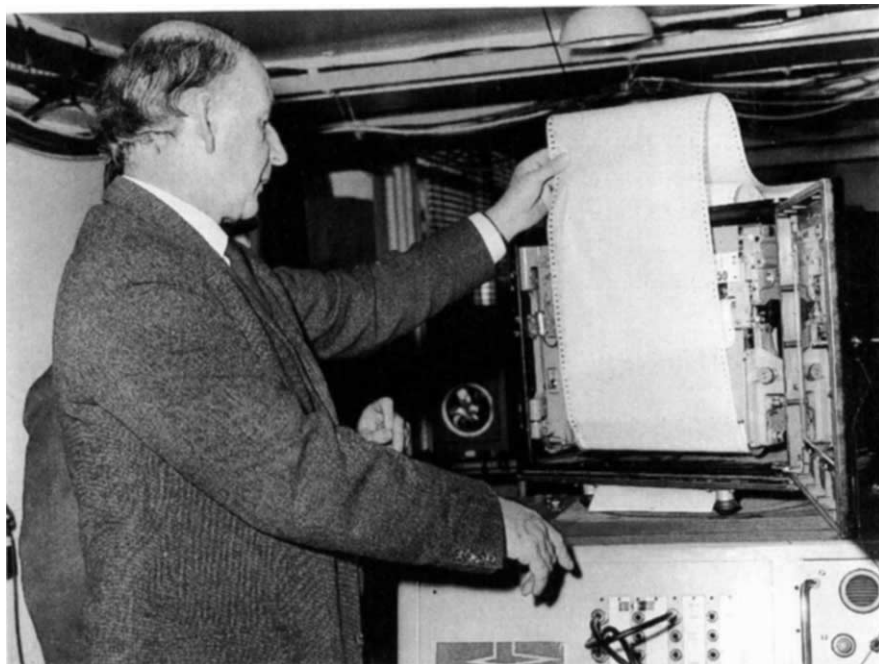
The work was done at the highest

pressure, spurred on by Churchill's adviser F. A. Lindemann (later Lord Cherwell) whose natural tendency to think offensively may have been enhanced by the prospect of a uniquely scientific aid in every bomber.

The resulting equipment, so brilliantly developed, ultimately received the code-

well's support for the offensive resulted in the maritime form of H₂S, as ASV Mark IV, being given second place. Fortunately, although the loss of H₂S over Germany should have alerted the Germans to its use at sea, ASV scored great successes against the U-boats. In fact Dönitz, the U-boat admiral, gave it the entire credit for the dramatic increase in U-boat sinkings in May 1943. Although the increase was no less due to Bletchley's contemporary 'breaking' of the naval four-wheel Enigma, which told us where U-boats were likely to be found, we were happy to leave Dönitz unenlightened.

H₂S had one feature that could to



Bernard Lovell: not only responsible for developing airborne radar equipment during the Second World War, but also one of the pioneers of radioastronomy.

name 'H₂S'. While Lovell is right in stating that it had an earlier designation of 'BN' (blind navigation), in the Cabinet Office and the Air Ministry I first heard it called 'TF', which might indiscreetly suggest 'town finding', just as 'AI' had been derived from 'air interception', and 'ASV' from 'air-to-surface vessel'. I therefore asked Lindemann to have the code-name changed, with results that I described in *Most Secret War* (Hamish Hamilton, London, 1978)*. Lovell, though, says "The device was never known as 'TF' as Jones states". Our memories also differ on how 'Home Sweet Home' originated.

Besides its prospective success as a bombing aid, H₂S offered great promise in finding U-boats in the Battle of the Atlantic. This deserved a higher priority than bombing German towns, but Cher-

some extent offset its benefit to the user: its transmissions could give warning of the bomber's approach to the opposing defences. Although the German exploitation of this fact was, as Lovell says, "hardly surprising", that was not how it appeared to those working at Malvern at the time. Despite every warning from Intelligence that the Germans had developed an appropriate receiver, Naxos, and were fitting it to their nightfighters, and that they had also set up a very efficient listening organization to exploit the radiations from H₂S and other equipment in our bombers, the Malvern researchers refused to believe us, unless one of them could re-examine all our evidence. Fortunately, their nominee was P. I. Dee, for whom we all had the greatest respect, and he was attached to my staff in August 1944 so that he could see the Enigma decrypts. A morning was enough to convince him that the Germans were detecting our bombers at ranges of some hundreds of miles, and

*Published by Coward, McCann and Geoghegan (New York) as *The Wizard War* in the United States.

even over our airfields during testing of the equipment before operational flights. Moreover, many of their night-fighters were equipped with Naxos receivers, which as well as indicating the general whereabouts of the bomber stream could, in the hands of skilled operators, enable fighters to home on to individual bombers. Lovell is inclined to discount the success of Naxos; but if it was not successful, why did the Germans build as many as 1,500 Naxos receivers and progressively fit 700 of them into night-fighters?

Whatever the qualifications about the success of H₂S in Bomber Command, its success at sea was unqualified and had been gained by the devoted efforts of the team led by Lovell, and of others who worked closely with it. Nor was it work and inspiration alone that had been demanded of them: two of Lovell's team and three EMI workers, one of them A. D. Blumlein, were killed when their trials aircraft crashed in June 1942.

In *Echoes of War*, Lovell has provided an authoritative chronicle of the history of one of the greatest devices to emerge in the Second World War, and sheds internal light on the scientific and military cultures that developed in a free society determined to resist the Nazi threat. The fact that H₂S was in operational use only 15 months after its conception testifies to what inspired determination can do. And the fact that those 15 months included not only the fatal crash but also the upheaval of transferring the team and families from Swanage to Malvern in the spartan conditions of war adds to the marvel of the achievement. □

R. V. Jones, from 1939 to 1945 head of scientific intelligence on Britain's Air Staff and scientific adviser to the British Secret Intelligence Service (M16), is at 8 Queen's Terrace, Aberdeen AB1 1XL, UK.

■ Nanotechnology will drastically alter the way in which we think about our lives, argue the authors of two recently published books. In *Unbounding The Future* (Morrow, \$23), K. E. Drexler, C. Paterson and G. Pergamit give a popular account of nanotechnology and what they believe might be its revolutionary consequences. More measured is *The Vital Machine* (Oxford University Press, £17.50), in which D. F. Channell takes a historical look at how various forms of technology have shaped the distinction we make between natural and artificial, alive and mechanical.

■ Cambridge University Press have reissued three introductions to particle physics. New in paperback is *The Experimental Foundations of Particle Physics* by R. N. Cahn and G. Goldhaber (£15.95, \$27.95). New editions are *The Ideas of Particle Physics* by G. D. Coughlan and J. E. Dodd (2nd edn; £35, \$59.95 (hbk); £13.95, \$27.95 (pbk)) and *Elementary Particles* by I. S. Hughes (3rd edn; £50, \$80 (hbk); £16.95, \$29.95 (pbk)).

Life and hard science

Walter Gratzer

Five Billion Vodka Bottles to the Moon: Tales of a Soviet Scientist. By Iosif Shklovsky. Translated by Mary and Harold Zirin. Norton: 1991. Pp. 268. \$19.95, £14.95

THE story goes that when Khrushchev turned on Stalin, by then safely dead, one of his diatribes was checked by a voice from the hall, demanding to know why, if all he said was true, he had not spoken up when it mattered. Khrushchev flushed dangerously and slowly scanned the audience with his bloodshot eyes. Would the questioner identify himself? There was a chilling silence and then Khrushchev spoke again: "Now you know". Among the scientific community it was pre-eminently Kapitsa who stood up to Beria and Stalin. But there were other brave men who made no deals with the system and retained their self-respect, among them the astrophysicist Iosif Shklovsky.

Shklovsky, it must be said, does not emerge from his discursive memoirs as an altogether attractive personality. "The style and flavour of Richard Feynman", promised by the blurb, is far removed from the querulous tone and leaden sarcasm here on display. But then while Feynman pursued his serene course to an early apotheosis, Shklovsky was being denied at every turn, thwarted in his own country of the recognition that he craved, harassed by the endemic antisemitism of the Soviet establishment and oppressed by a brainless and pusillanimous bureaucracy. He fought the system with tireless tenacity until the casual incompetence of the doctors at the Soviet Academy of Sciences hospital put an end to him. It was Dumas fils who said that rogues are preferable to imbeciles because they sometimes take a rest.

Despite the rancorous descent, Shklovsky's story is a valuable and often absorbing record of the time. Science — hard enough, God knows, even when everything is in your favour and everyone is on your side — ought to have been all but impossible in the conditions with which Shklovsky had to contend. He chronicles, for instance, the fall of the celebrated astronomer. Numerov, who was denounced to the Communist Party by an unsuccessful graduate student out for revenge. Numerov under torture signed a fabricated confession, admitting his part in an anti-Soviet conspiracy and implicating two dozen of his colleagues. The contagion spread and in the end some 80 of the best astronomers in the country were imprisoned and most of them lost their lives.

The absurdities to which fear and bureaucratic paralysis could lead are

illustrated in many chilling anecdotes, often rich in black humour. In the Karpov Physical Chemistry Institute, an order was posted containing, over the signature of the director, the following article: "4. Associate of the Electrochemical Laboratory Comrade Morokhov is hereby reprimanded for applying to the Karpov Institute a word that in the language of poorly educated people is used to denote a house of prostitution." A month later another order appeared, rescinding article 4, and thus, as Shklovsky notes, endorsing Comrade Morokhov's thesis. And here is an example of Soviet medicine at work: Shklovsky, in the academy hospital after his first heart attack, takes his pulse to save the doctor the trouble. "Seventy-three", he announces. The doctor, a well-dressed young woman, and therefore surmisedly the daughter of an academy panjandrum, replies that this is impossible, the heartbeat is always an even number: her method, it turns out, is to count for half a minute and then double the result.

Shklovsky comes across then as cross-grained, unyielding and valiant, a pain to the *apparat* and often probably to his friends, not given to self-doubt and sustained by his high opinion of his abilities and accomplishments, which he is not bashful about enumerating. And who should say his pride misplaced? His story is marvellously translated and edited by Mary and Harold Zirin, the latter himself a professor of astronomy at Caltech and director of its observatory. It is their insiders' knowledge of the man, his science and the social and professional milieu in which he functioned that make this book rewarding. Here is Shklovsky, weak with hunger on a farm in the remote provinces, stealing potatoes; he becomes proficient at gathering them, he says, by the method of Mamlakat. The editorial footnote explains: "Mamlakat Nakhangova was a little Uzbek girl, who became a heroine of labour for picking cotton with both hands." Now where else could one have learned that? □

Walter Gratzer is in the MRC Cell Biophysics Unit, King's College London, 26–29 Drury Lane, London WC2B 5RL, UK.

Leigh Royden's review of *The Solid Earth* (*Nature* **354**, 30; 1991) inadvertently rendered the author of the book, Mary Fowler, male. The mistake was made in the editorial office of *Nature*.

Sacred insects

Malcolm Coe

Dung Beetle Ecology. Edited by Ilkka Hanski and Yves Cambefort. Princeton University Press: 1991. Pp. 481. \$60. £43.

THE apparently ceaseless energy with which the scarab beetle rolls its ball of dung has fascinated humans for at least 3,500 years. Ancient Egyptians saw in this activity an image of the Sun moving across the heavens, thus giving the humble insect its divine attributes. Oddly

fossil record about 250 million years ago, which leads one to wonder whether the dung of large herbivorous dinosaurs led to the evolution of the first true dung beetles. It seems, though, that the main radiation of the modern species must have coincided with the radiation of herbivorous mammals much more recently. The shift from feeding on abundant but low-quality rotting vegetation to scarce and patchy high-quality resources was important in the evolution of the habits exhibited by today's dung beetles: rollers, which remove balls of dung from the site of deposition to bury them elsewhere for the purposes of feeding (maturation) or breeding; tunnellers,

Labours lost

Stuart Sutherland

The Science of Love: Understanding Love and Its Effects on Mind and Body. By Anthony Walsh. Prometheus: 1991. Pp. 276. \$22.95, £14.50.

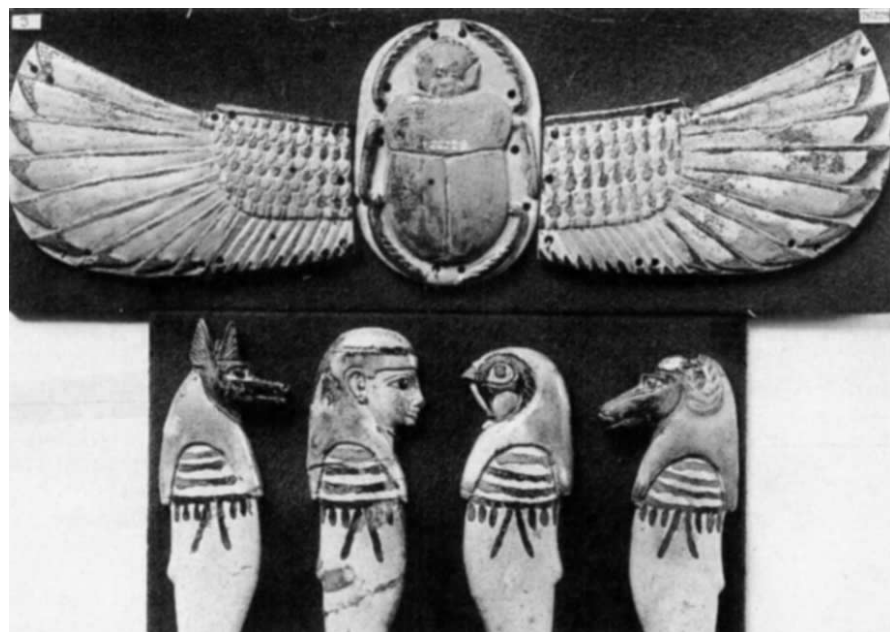
AN authoritative dictionary of psychology defined love as "a form of mental illness not yet recognised in the standard diagnostic manuals". Walsh will have none of such cynicism: indeed, he writes, "Only when we are loved and can give love in return do we feel whole . . . we yearn to be connected. Love beautifies our lives, it empowers our being, it ennobles us, it enriches us in every way, and it imbues our minds and hearts with a sense of the fullness of life." In short, Walsh is an enthusiast. He strives to demonstrate that love is a good thing and where possible to exhibit its physiological and psychological causes and consequences.

He argues that the absence of love, particularly in childhood, has disastrous effects. Certainly, Harlow's notorious infant monkeys, which were fed from an artificial metal mother, showed severe disabilities, but they made a considerable recovery later when allowed to consort with their peers. Although it is plausible to suppose that infants who are loved become happier adults than those who are not, there is precious little evidence to support this thesis. Walsh cites a few carefully selected findings that favour his case, but most studies, including some by John Bowlby, at one time the chief supporter of the importance of being loved as a child, suggest that deprived children usually make a complete emotional and intellectual recovery if they are moved to a better environment.

Walsh is on safer ground in claiming that bereavement and separation render people prone to all manner of illnesses and he sketches what is known of the harmful effects of unhappiness on the immune system. But as any hermit knows, there is a difference between not being loved and ceasing to be loved. Indeed, many psychiatrists believe that the effects of bereavement are greater in loveless marriages than in loving ones, an issue for which there is no good evidence. It is, however, known that any form of stress, such as that imposed by unemployment, moving house or even gaining promotion, can have the same effects on health as the loss of love.

In an attempt to justify the title of his book, Walsh strives to make connections between love and neurophysiology. In one of many wild speculations, he argues that "lack of love in infancy leads to

The Mansell Collection



The winged scarab beetle as revered by the ancient Egyptians.

enough, in most of the many representations of the scarab, it is shown with the ball (or Sun) between its forefeet, while in reality the dung is rolled with the hindfeet. Cambefort has recently proposed that the attitudes of the ancient Egyptians to the scarab combined three important elements of these people's lives: the Sun, soil and cattle. In addition, signs of renewal may have been provided by the emergence of the beetle from its pupa in the brood ball, and the pyramids may have represented a stylized dung pat.

The editors of this volume have brought together excellent summaries of our current knowledge about these intriguing creatures' ecology and the evolution of their coprophagous lifestyle. It seems fairly clear that this lifestyle evolved from that of their saprophagous ancestors; even today there are many dung-beetle species and close relatives that feed on rotting vegetation rather than the digested plant litter of herbivorous mammals. The ancestors of these dung-feeding beetles first appear in the

which construct feeding or breeding burrows below the dung pile; and dwellers (generally small), which complete their feeding or breeding within the dung pile.

The great French entomologist Henri Fabre was the first to describe in detail the remarkable habits of dung beetles, especially those that show a high degree of brood care in raising their young within dung balls, which are often buried underground in specially constructed brood chambers.

Today, when many habitats and their accompanying herbivore faunas are being drastically altered or even exterminated by human activity, a detailed understanding of the ecological role and diversity of these diminutive recyclers of nutrients is vital. The book is directed primarily at those concerned with this ecology, but there is a great deal here for the general reader interested in "the natural history of the unmentionable". □

Malcolm Coe is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

inadequately developed pathways to the pleasure centres of the brain [and] a later inability to fully experience love and pleasure". He claims the psychopath has an underactive autonomic system and the schizophrenic an overactive one: in both cases the cause is alleged to be lack of love. He does not mention the harmful effects on schizophrenics of having love too strongly expressed, as demonstrated by J. Leff and J. K. Wing. And he appears to think (in common with many neurophysiologists) that finding a correlation between a neurotransmitter and behaviour solves everything, whereas the interesting question is how its effects are mediated by the nervous system.

Walsh's book is an extraordinary potpourri of ideas drawn from many different disciplines, including sociology, neurophysiology, ethology and sociobiology. Much of it fails to inspire confidence. For example, he believes in group selection rather than selection for the individual gene; he revives Maclean's discredited theory of the triune brain; he thinks infants have to learn to walk; and he writes "altruism takes place in the limbic system" though he fails to tell us whether that system is kind to the cortex or to the cerebellum. He even manages to resuscitate that old fraud Teilhard de Chardin who saw love as a "universal

psychic energy" that would reach "Point Omega" somewhere up in the "noosphere".

Moreover, many of his arguments are topsy turvy. For example, he states that maternal love is stronger than paternal *despite* the inconvenience and pain of pregnancy and parturition. Surely, it is, at least in part, the mother's suffering that makes her value the end product so much, for it is well established that the more a person suffers to produce something, the more he or she values it.

Among the gush and nonsense, there appears the occasional interesting finding or remark, though even these often arouse scepticism. Is it really true that from an examination of placental DNA in a variety of races one can conclude that mankind is descended from one woman who lived 200,000 years ago? Walsh is more interesting on the history of marriage than on the role of oxytocin in promoting maternal love. One is driven to the conclusion that love is not yet a suitable subject for scientific study. Indeed, Ovid knew more about "the chemistry of love" than Walsh, who confines himself to the literal meaning of that expression. □

Stuart Sutherland is in the Department of Experimental Psychology, University of Sussex, Falmer, Brighton BN1 9QG, UK.

Good coverage

William A. Steele

Dynamics of Gas-Surface Interactions. Edited by C. T. Rettner and M. N. R. Ashfold. *Royal Society of Chemistry: 1991. Pp. 366. \$97.50.*

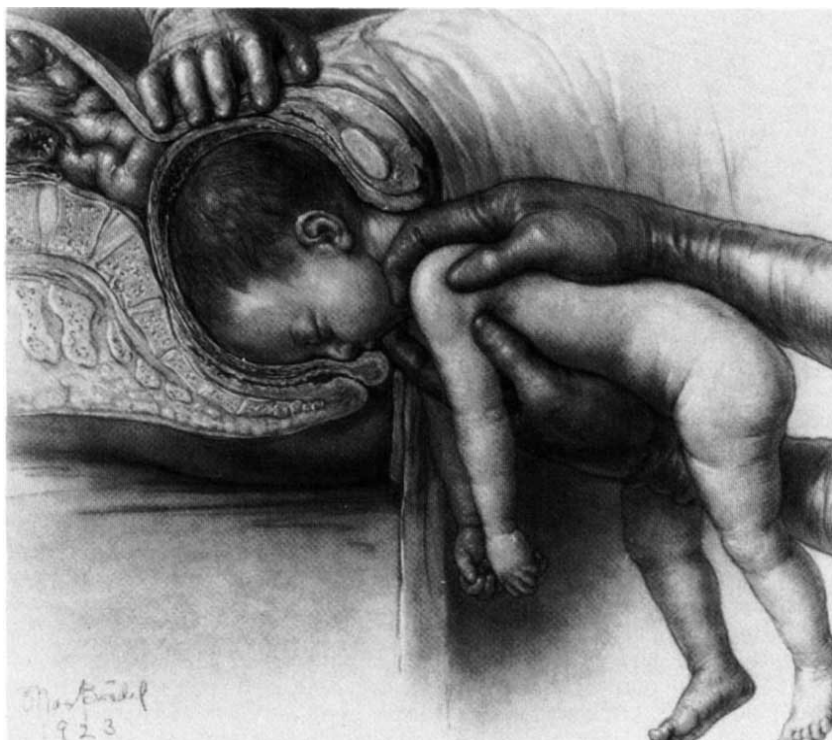
GAS-SURFACE interactions crop up in many systems, from the physical interactions of inert gases with the heterogeneous surfaces of solids with large surface areas, to the chemical interactions of gases with clean, physically perfect crystalline adsorbents. Over the past few decades, there has been a rapid advance in understanding of these systems. The authors of this book give excellent, well-written summaries of the present state of knowledge, primarily of the chemical bonding and dynamics of simple molecules at coverages of less than a monolayer on clean, well-characterized solid surfaces (mostly metallic).

Physical interactions and their consequences are seldom discussed other than in a chapter on energy transfer in gas-solid collisions. Most of the descriptions of chemical bonding at surfaces and of selected experimental probes of this bonding are limited to coverages sufficiently low that isolated adsorbed species are the focus of discussion. The exceptions are chapters on the chemical reactions of adsorbed species and on the kinetics of thermal desorption.

Two articles are explicitly devoted to descriptions of specific experimental techniques; namely, electron-stimulated desorption and photochemistry in the adsorbed state. The successful interpretation of desorption measurements requires a reasonably detailed model of the process, and this aspect of the problem is thoroughly discussed. There is also a chapter on theoretical progress in describing energy transfer in particle-surface collisions, as well as two chapters on the theory of dynamical chemisorption of simple gases on metal surfaces.

The problems covered in the book lie at the heart of surface chemistry, but much more theoretical insight and computing power will be required to crack them. To their credit, the authors are frank in their discussions of the difficulties as well as achievements. But if this is one of the book's strengths, then it is also a weakness, for the volume, although authoritative for the time being, is likely to become rapidly outdated. □

William A. Steele is in the Department of Chemistry, 152 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.



Max Brödel (1870–1941) is widely regarded as the father of medical illustration. Founding director of the pioneering academic department devoted to the subject at Johns Hopkins University, he collaborated with many distinguished physicians to provide them with teaching drawings such as the one shown here of a breech birth. Many more of these magnificent pictures are reproduced in *Max Brödel: The Man Who Put Art Into Medicine*, a superb biography by R. W. Crosby and J. Cody. Published by Springer, price £35, \$49.

Quantum chaos

Roderick V. Jensen

What are the special properties of strongly perturbed quantum systems when the corresponding classical mechanical systems exhibit chaotic behaviour? This is the central question of quantum chaos. Theoretical and experimental studies of highly excited atoms in strong magnetic fields and intense microwave fields have revealed surprising new physical phenomena with far-reaching applications.

"The classification of the constituents of a chaos, nothing less is here essayed."

Herman Melville
Moby Dick, 1851

IN Melville's time, 'chaos' referred to a state of disorder or confusion. In the past 15 years, the word has been endowed with a new meaning in the lexicon of science and mathematics. Since May's seminal review article¹ in 1976, "*Simple mathematical models with very complicated dynamics*", chaos has emerged as a technical term for the irregular, unpredictable and apparently random behaviour of a wide variety of deterministic dynamical systems such as fluctuating biological populations, cardiac arrhythmias, oscillating electrical circuits, vibrating structures, turbulent fluids and particle orbits in accelerators, plasma fusion devices and the Solar System²⁻⁴. In turn, this new concept of 'deterministic randomness'⁵ has given birth to a rapidly developing interdisciplinary field of research called nonlinear dynamics⁶.

My goal here is less ambitious than Melville's stated aim in his well-known treatise on 'chaos'. Although the many new concepts and research methods in nonlinear dynamics and chaos have found widespread applications in the biological and social sciences as well as in the physical sciences and engineering, I will focus here on the classical and quantum mechanical description of strongly coupled and strongly perturbed nonlinear systems in physics. Specifically, I will address the profound and often controversial problem of 'quantum chaos', the question of how quantum systems behave when the corresponding classical systems are chaotic. Extensive theoretical modelling coupled with experimental measurements on physical systems has uncovered a wealth of new physical phenomena at this exciting frontier between the microscopic and the macroscopic world, where the classical theory is chaotic and the quantum theory can be very complicated.

Classical chaos

According to Newton, the behaviour of a mass on the end of spring or a planet orbiting the Sun should be completely described by the differential equations of motion of classical mechanics⁷. If we specify the initial values of the positions and momenta, then the deterministic equations uniquely specify the motion for all times in the future (and the past). But except in rare examples, usually emphasized in physics textbooks, these equations do not have simple analytical solutions.

In general the classical equations of motion are nonintegrable. Although the evolution is deterministic, in most cases we cannot find it without the help of numerical calculations. Worse, numerical studies of the nonlinear classical equations of motion for many model problems, such as two or more coupled, anharmonic oscillators⁸, or asteroid orbits periodically perturbed by the gravitational pulls of nearby planets⁹, have led to the realization that the time evolution can show 'extreme sensitivity to initial conditions'. Uncertainties or errors in the initial conditions can grow at an exponential rate which is intrinsic, not an artefact of any numerical approximations. Remarkably, this property of nonlinear differential equations was first pointed

out by Poincaré¹⁰ at the turn of the century for the classical three-body problem, long before the advent of modern digital computers.

The local instability of motion that leads to this 'extreme sensitivity to initial conditions' is the defining property of chaos in classical systems. The laplacian ideal of perfect predictability in these deterministic systems can only be realized if the initial conditions are known and specified with infinite precision. Any errors grow rapidly, preventing long-time prediction and resulting in behaviour that appears irregular and random⁴.

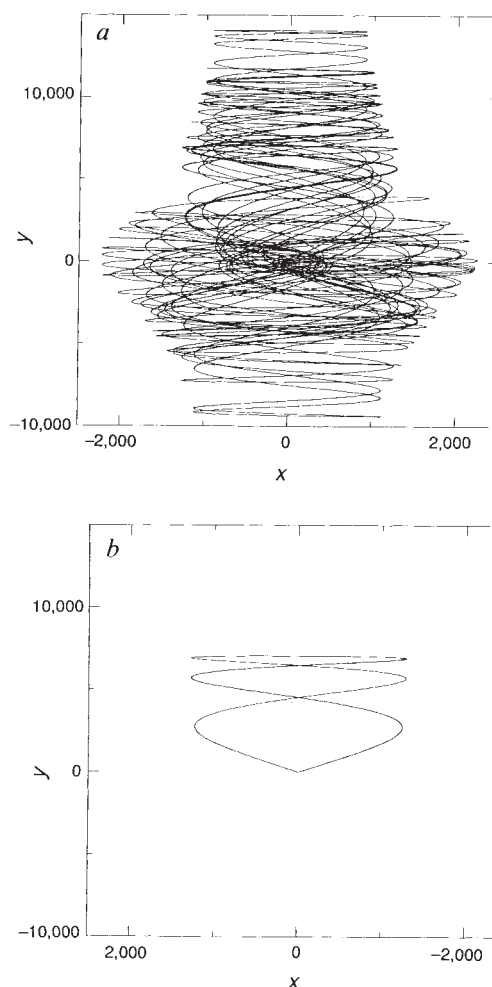


FIG. 1 a, A typical chaotic, classical orbit of an electron moving in the two-dimensional potential in equation (1). b, For special initial conditions⁴⁶ the trajectory can be periodic: the particle moves away from the nucleus at the origin, oscillates because of the force of the magnetic field, and returns to the nucleus where it is reflected backward along its previous path. For strong magnetic fields, however, these regular, periodic orbits are highly unstable. If we displace the trajectory by as little as one part in 10^6 , then we get the highly irregular orbit shown in a.

To illustrate this concept of deterministic randomness, Fig. 1 shows plots of two different classical trajectories for a particle moving without friction in a two-dimensional potential well¹¹

$$V(x, y) = -1/\sqrt{x^2 + y^2} + \frac{B^2}{8} x^2 \quad (1)$$

which corresponds to the effective potential experienced by an electron moving in an attractive Coulomb field and a constant magnetic field, B . If the potential consisted only of the first term representing the Coulomb field or the second term corresponding to a simple harmonic oscillator, then the analytical solutions of the equations of motion could be found in many classical physics texts⁷. But this simple combination of forces often yields irregular, unpredictable, apparently random trajectories such as that shown in Fig. 1a. For a slightly different initial condition, corresponding to a change of the starting position by one part in a million, the resulting trajectory (Fig. 1b) is completely different. In this case it lies close to one of the infinitely many (but still relatively rare) unstable periodic orbits for this dynamical system.

The problem of quantum chaos

Bohr's correspondence principle¹², which requires that quantum and classical behaviour should coincide for macroscopic systems, might suggest that some vestiges of classical chaos should persist in the quantum theory. For bounded systems, however, neither the quantum wavefunctions nor any observable quantities can show the extreme sensitivity to initial conditions that defines classical chaos. The quantum mechanical energy spectrum is discrete, and the solutions of the Schrödinger equation restrict the quantum dynamics to quasiperiodic behaviour, whereas the corresponding classical dynamics can be fully chaotic.

As a consequence it does not seem possible to define quantum chaos as a property of a quantum system that corresponds to chaos in classical mechanics. Nevertheless, the term quantum chaos is used to refer to the study and description of the properties of quantum systems for which the corresponding classical hamiltonian system is chaotic. In an effort to avoid confusion between the definition of an attribute of a physical system and a field of study, Berry¹³ has introduced the term 'quantum chaology' to be synonymous with this usage of 'quantum chaos'.

It may emerge from these studies of quantum chaology that real quantum systems can exhibit true chaos. Indeed, some peculiar theoretical models of unbounded quantum systems (with continuous energy spectra) have been shown to be capable of chaotic dynamics, but these models suffer from unphysical properties such as unbounded, exponential growth of energy and momentum¹⁴. For realistic, few-body, quantum systems, such as atoms and molecules, it seems that radical changes in the present quantum theory are required for chaotic behaviour⁵.

Despite debate over the meaning of the term 'quantum chaos', nearly everyone agrees on the important problems¹⁵. What is the nature of the classical-quantum correspondence principle for classically chaotic systems? What are the characteristic properties of the energy levels and eigenfunctions for a system of strongly coupled nonlinear oscillators, such as a highly excited polyatomic molecule or a simple hydrogen atom in a strong magnetic field? How does the quantum wavefunction evolve for strongly perturbed nonlinear oscillators such as atoms or molecules in intense electromagnetic fields?

These problems represent an exciting frontier for modern physics and chemistry. Pioneering experiments on highly excited Rydberg atoms in strong magnetic static fields have been made by groups at the University of Bielefeld¹⁶⁻¹⁹ and Massachusetts Institute of Technology²⁰⁻²² and in intense microwave fields at the University of Pittsburgh^{23,24} and the State University of New York at Stony Brook²⁵⁻²⁷. The classical descriptions of these highly excited electron orbits are chaotic when perturbed by

sufficiently strong magnetic and electromagnetic fields. For example, when the Lorentz force on an electron moving in a strong magnetic field competes with the Coulomb binding field^{11,28-30}, then the typical classical electron trajectory resembles that shown in Fig. 1a. Similarly, when the oscillating force due to an applied microwave field becomes comparable to the Coulomb binding field³¹, a stroboscopic picture of the classical electron motion in action-angle phase space shows a transition from mostly regular behaviour (Fig. 2a) to widespread chaos (Fig. 2b). Although the solutions of the Schrödinger equations for these systems cannot satisfy the strict definition of classical chaos, experimental and associated theoretical studies of the quantum chaology for these simple atomic systems have nevertheless revealed a rich variety of new physical phenomena.

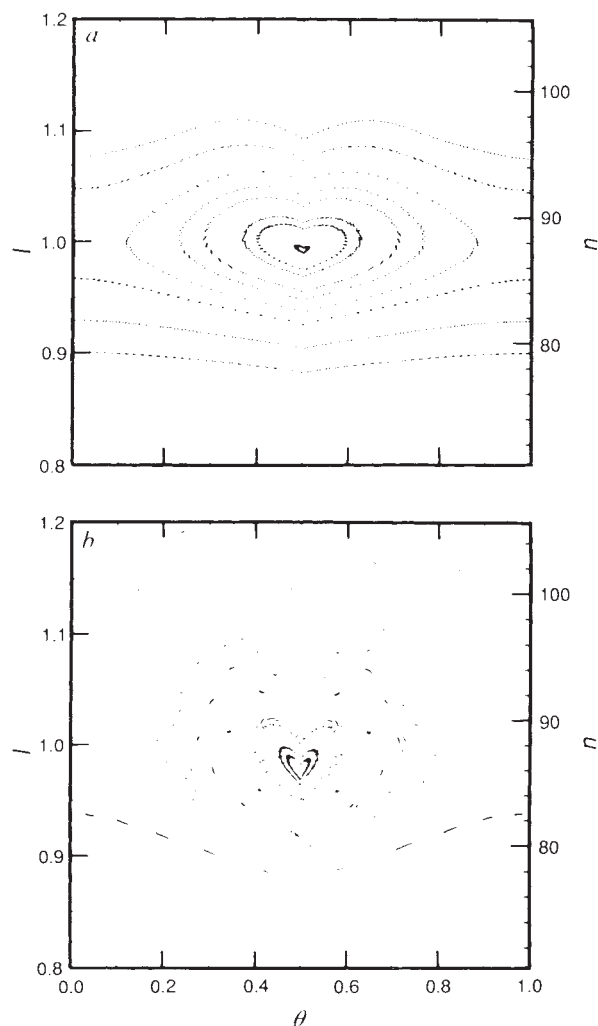


FIG. 2 Poincaré surfaces of section provide a revealing stroboscopic picture of the classical dynamics for a bound electron in a one-dimensional hydrogen atom perturbed by a very intense microwave field. Instead of using the usual position and momentum variables, I have transformed to the more convenient action-angle variables (I, θ) where the action, I , is proportional to the principal quantum number n by the Bohr-Sommerfeld quantization, and the angle variable, θ , is conveniently measured in units of radians/ 2π . The numerically calculated locations are shown for several classical orbits plotted once every period of the oscillating perturbation. In a, the perturbing field is below the threshold for the onset of chaotic ionization and the stroboscopic traces of the classical orbits lie along smooth, regular curves for initial conditions corresponding to highly excited hydrogen atoms with principal quantum numbers between $n=80$ and 100. When the perturbing field is increased above the threshold for chaotic ionization (b), many initial conditions lead to highly irregular, chaotic orbits which rapidly ionize. (Reprinted with permission from ref. 33.)

For example, measurements and calculations of the spectra of hydrogen atoms in strong static magnetic fields^{28–30} have revealed relationships between the statistics of the energy level spacings and the predictions of the random matrix theory of nuclear physics³². In addition, detailed investigations of the electron dynamics in a hydrogen atom exposed to strong electromagnetic fields have shown how the quantum dynamics can mimic the classical chaos for limited times^{33–34} and how quantum interference effects can suppress the chaotic classical dynamics by localizing the wavefunction^{35–37} through a mechanism related to Anderson localization in solid-state physics.

Moreover, theory and experiment for both systems have exposed the importance of 'scarred' wavefunctions which are highly correlated with unstable classical periodic orbits³⁸. Even though the classical description of the electron motion in a highly excited hydrogen atom in an intense microwave field or in a strong magnetic field may be fully chaotic, many quantum wavefunctions nevertheless seem to cling to the remnants of regularity represented by these unstable periodic orbits.

Highly excited atoms in strong magnetic fields

The Zeeman splitting of atomic energy levels by magnetic fields³⁹ has traditionally been a proving ground for new theoretical ideas and approximations in the development of quantum mechanics. For weak and moderate magnetic fields, the splitting is well described by quantum perturbation theory⁴⁰. For example, in moderate magnetic fields the electron energy levels in hydrogen atoms have the regular sequence (Fig. 3a) that is predicted by the theory of the quadratic Zeeman effect⁴¹. In large magnetic fields, however, where the magnetic forces on the electron become comparable to the Coulomb binding fields, the conventional tools of quantum perturbation theory break down and the energy levels become 'irregular', as shown in Fig. 3b. This transition from regular to irregular spectra, as the corresponding classical system goes from regular to chaotic, was first conjectured by Percival⁴² and Zaslavsky⁴³ in the 1970s, although this conclusion was presaged by Einstein⁴⁴ in 1917 when he remarked on the difficulty of assigning quantum numbers to the energy levels in nonintegrable systems.

For ground-state atoms this transition occurs for astronomically large fields ($\sim 10^4$ Tesla) which can only be realized in white dwarf and neutron stars. But as the Coulomb binding fields decrease rapidly as $1/n^4$ with increasing principal quantum number n , this new parameter regime where the classical mechanics is chaotic and the quantum spectrum is complex can be readily achieved using highly excited atoms with principal quantum numbers ranging from $n = 30$ to 40 in laboratory fields of 4–6 T. As a consequence, the energy spectrum of highly excited hydrogen atoms has emerged as an important model problem in this new field of quantum chaos.

Because of the complexity of the irregular energy spectrum for hydrogen atoms in strong magnetic fields (Fig. 3b) and for other model quantum systems that are classically chaotic, the spectra were first analysed with statistical methods that had been previously developed to characterize the complex, irregular energy-level spectra of excited nuclei^{28–30}. Because the precise coupling between strongly interacting nucleons is not known and is presumed to be complicated, it was simply assumed that the statistical properties of the nuclear energy-level spacings would be indistinguishable from the distribution of eigenvalues of large matrices with randomly chosen matrix elements³². This random matrix theory predicts that the probability distributions for the spacings, s , of adjacent energy levels should be well approximated by the Wigner–Dyson distribution $P(s) \approx s \exp(-s^2)$, which provides an excellent description of the spacings of the measured energy levels in excited nuclei³².

Although there are no unknown or random elements in the hamiltonians for these classically chaotic systems, the statistics of the energy-level spacings for these systems also agreed remarkably well with the specific predictions of the random

matrix theory. This striking result suggests that the quantum dynamics of the strongly interacting nucleons in excited nuclei may share many features with the much simpler hydrogen atom in a strong magnetic field³². Nuclear physicists have been greatly interested by these results, and some have even proposed that these statistical properties of the energy distributions should form the definition of quantum chaos³². But this statistical 'structure' in energy-level spacings is only one of the many interesting properties of quantum systems that are classically chaotic.

The first experiment on highly excited atoms in strong magnetic fields to indicate that there was even more interesting structure in the spectrum than that described by the statistical theories were done by Garton and Tompkins⁴⁵ in 1969. They measured the photoabsorption spectrum by exposing barium atoms in strong magnetic fields to different wavelengths of light. When the photon energy of the light matched the excitation energy for a highly excited state, the production of excited states could be detected. Although the distribution of energy levels was expected to resemble well-stirred spaghetti (like the pattern shown in Fig. 3b), these experiments produced one of the first surprises of the quantum description of classically chaotic

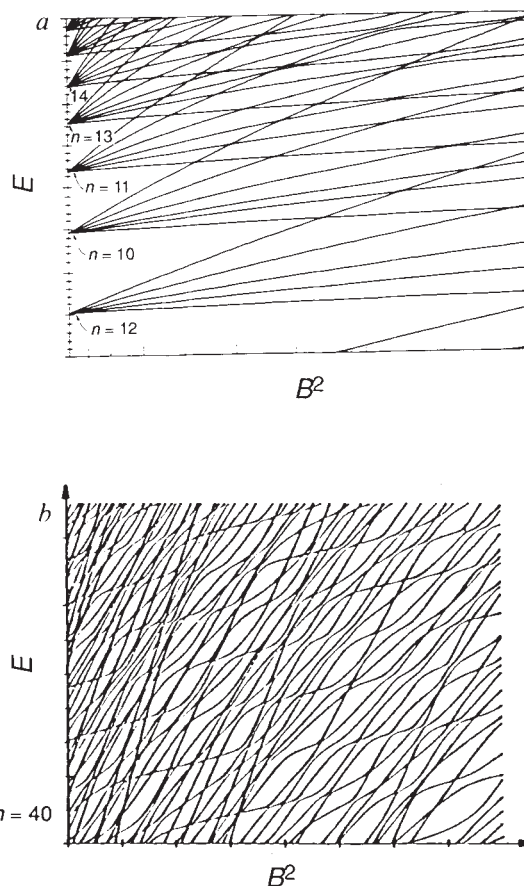


FIG. 3 The discrete, quantum-mechanical energy levels of a hydrogen atom in an external magnetic show a well studied symptom of 'quantum chaos'^{28–30}. When the magnetic forces on the electron are much weaker than the Coulomb forces (a), the low-lying energy levels of the hydrogen atom with $n=10-16$ in static magnetic fields have a regular spacing which increases with B^2 as predicted by the theory of the quadratic Zeeman effect. (This figure was provided by D. Kleppner.) b, For highly excited atoms with $n \geq 40$, magnetic fields of 5–6 T strongly perturb the electron energy levels. The classical dynamics in the combined Coulomb and magnetic fields is chaotic, and the calculated and measured spectra show irregular patterns resembling well stirred spaghetti. (Reprinted with permission from ref. 74.)

systems by exhibiting regular modulations of the photoabsorption cross section as a function of energy⁴⁵.

Because of the enormous number of quantum levels spanned by these early experiments, it was difficult to resolve the fine structure of the spectrum, let alone individual quantum levels. Higher-resolution measurements of photoabsorption in strong magnetic fields of highly excited hydrogen atoms¹⁶⁻¹⁹ and of highly excited lithium atoms²⁰⁻²² provided a detailed picture of the rich and intricate structure of the spectrum. For example, Fig. 4a shows the measured photoabsorption spectrum for hydrogen atoms in a strong, 5.96-T magnetic field. The graph shows the laser-excitation probability from an $n=2p$ state to high-lying Rydberg states with principal quantum numbers $n >$

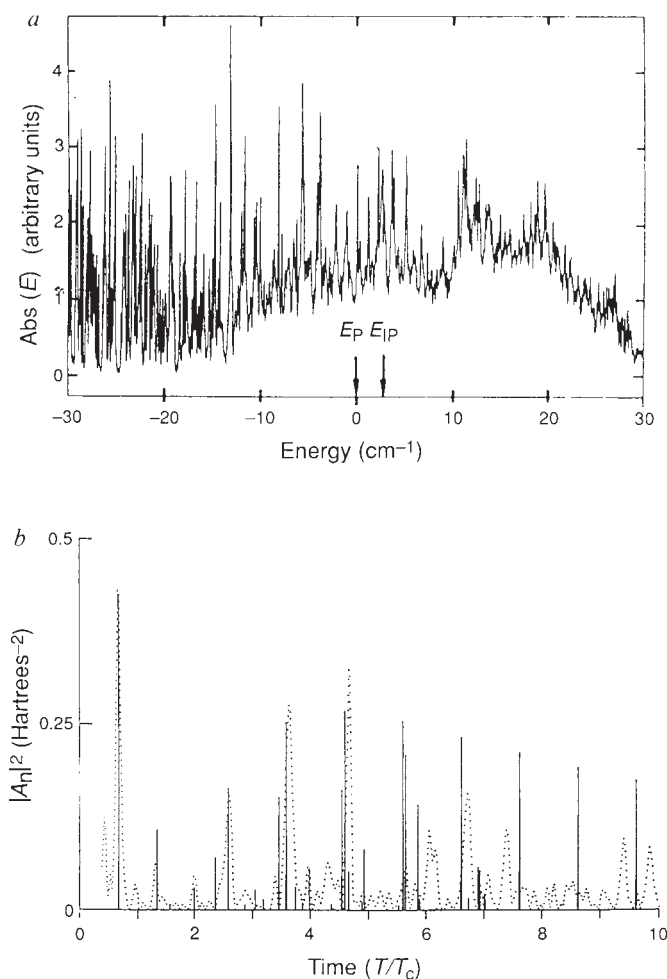


FIG. 4 a, The photoabsorption spectrum for a hydrogen atom in a strong, 5.96-T magnetic field as a function of laser energy for transitions near the ionization potential E_{IP} , which is shifted slightly above 0 in the presence of the strong magnetic field. If the experiment had unlimited resolution, then the spectrum for $E < E_{IP}$ would consist of an infinite sequence of sharp peaks located at the transition energies in the strongly perturbed atom and the heights of the peaks would provide a measure of the transition probability. This 'stick' spectrum is smoothed because of the finite experimental resolution and the natural line widths of the ionizing states for $E > E_{IP}$. Although this complex spectrum seems to have little structure, the reader may still discern a few Garton-Tompkins oscillations¹⁶ with period $\sim 10 \text{ cm}^{-1}$ on the right side of the figure. (Reprinted with permission from ref. 18.) b, The Fourier transform of the photoabsorption spectrum in a (dotted curve), compared with semiclassical theoretical predictions based on the contributions from short periodic (or closed) orbits in the classical description of a highly excited hydrogen atom in a strong magnetic field. The locations of the peaks in the Fourier transform, which represent regular modulations of the quantum photoabsorption spectrum, correspond to the periods of these short orbits measured in units of the classical gyro-period of a free electron in the static magnetic field, T_c . (Reprinted with permission from ref. 46.)

60 for laser photon energies $E < E_{IP}$ and into free-electron resonance states for $E > E_{IP}$, where E_{IP} is the ionization potential. This complex spectrum, representing the contributions of millions of individual energy levels, does not seem obviously regular. If, however, we plot the Fourier transform of this photoabsorption spectrum as a function of energy, then we find distinct peaks as a function of time, as shown by the dotted curve in Fig. 4b. For example, the original Garton-Tompkins oscillations correspond to a characteristic time, $2/3$ the classical gyro-period of a free electron in the magnetic field, which gives rise to the largest peak in Fig. 4b. Many more peaks are clearly visible.

Extensive theoretical work^{18,19,28-30} has now shown that the characteristic times for these regular oscillations in the 'irregular' spectrum correspond to the periods of short, usually unstable, periodic orbits in the classical description of electron motion in combined magnetic and Coulomb fields, such as that shown in Fig. 1b. The solid lines in Fig. 4b indicate the results of theoretical calculations of the photoabsorption spectrum based on the contributions of interfering wavefunctions that propagate along the directions of these unstable classical orbits⁴⁶. Moreover, additional high-resolution measurements of photoabsorption by lithium atoms in strong magnetic fields^{21,22} show that these regular features in the spectrum associated with unstable classical periodic orbits clearly persist for positive energies where the typical classical electron trajectories are not only highly unstable but are not even bound.

Quantum 'scars' of classical orbits

Although the solutions of the time-dependent Schrödinger equation for this bounded quantum system cannot exhibit the local instability of chaotic classical trajectories, like the one shown in Fig. 1a, the eigenfunctions and eigenvalues nevertheless seem to be strongly correlated with the more regular classical periodic orbits, like the one in Fig. 1b. This remarkable property of the quantum description of classically chaotic systems was strikingly demonstrated in a simple model problem known as the 'stadium billiard'³⁸. Figure 5a shows a typical classical trajectory of a particle moving without friction on a billiard table shaped like a stadium. This model system can be rigorously proven to be fully chaotic. Although there is a dense set of periodic trajectories, they are all unstable.

The quantum description is equivalent to solving for the electromagnetic modes in a cavity or the vibrations of a drum-head of this shape⁴⁷. In spite of the irregular classical dynamics, many of the quantum eigenfunctions (or normal modes of vibration) have surprising structure with high amplitude along short, unstable periodic orbits. For example, Fig. 5b shows a contour map of the probability amplitude for a highly excited eigenstate of the stadium. Heller³⁸, who first observed these structured eigenstates in numerical calculations of the eigenfunctions of the stadium billiard, described them as having 'scars' of the classical periodic orbits, and provided theoretical arguments for the constructive quantum interference responsible for their appearance.

These 'scarred' wavefunctions were subsequently found in other classically chaotic quantum systems: they exist in the model systems corresponding to smooth molecular potentials⁴⁸, in numerically calculated wavefunctions for hydrogen atoms in strong magnetic fields⁴⁹ and in experimental measurements of the eigenmodes of microwave cavities with walls shaped like classically chaotic billiards⁵⁰. The scars seem to be ubiquitous in quantum descriptions of classically chaotic systems, and this has important implications. In particular, the existence of these highly structured wavefunctions suggests that a reinterpretation of the correspondence principle is necessary, at least in this intermediate regime between the microscopic quantum and macroscopic classical worlds. For example, because of this 'localization' of some of the quantum wavefunctions to the vicinity of classical periodic orbits, statistical theories that are

applicable to the ergodic, classical systems may fail for the quantum.

Theoretical research into the mathematical connection between the classical periodic orbits and the spectral oscillations and scarred wavefunctions using the elaborate, semiclassical periodic orbit theory of Gutzwiller⁵¹, which is based on the Feynman path-integral formulation⁵² of quantum mechanics, is now beginning to bear much fruit. For example, using the unstable periodic orbits of the highly chaotic motion of two classical electrons orbiting the helium nucleus, Ezra *et al.* have succeeded in doing a semiclassical calculation of the energy levels of doubly excited states of helium atoms⁵³. This striking illustration of the classical-quantum correspondence for classically chaotic systems finally provides the solution of a 70-year-old problem⁵⁴ which had originally led to the demise of the old Bohr-Sommerfeld quantum theory in the 1920s.

Microwave ionization of excited atoms

Einstein's 1905 theory of the photoelectric effect⁵⁵ stated that the ionization of an atom exposed to electromagnetic radiation

requires the absorption of a photon with the right frequency (or energy) to make the transition from a bound state to the continuum. As a consequence, photoionization is expected to depend very sensitively on the radiation frequency but not on the intensity. But in 1974, Bayfield and Koch²³ reported a remarkable result from their experiments on microwave ionization of highly excited hydrogen atoms with principal quantum numbers near $n = 66$. They found that the ionization depended strongly on the intensity of the radiation and only weakly on the frequency. When the microwave fields were below a critical threshold, no ionization occurred. When the microwave fields exceeded a critical threshold, corresponding to 10–20% of the Coulomb binding field for these highly excited states, the atoms were rapidly ionized.

As the photon energy was only $\sim 1\%$ of the energy required to ionize these atoms, this novel ionization mechanism was presumably a high-order multiphoton process (one hundred photons or more). A quantum mechanical, time-dependent perturbative treatment of this system was not feasible because both the magnitude of the perturbation and the numbers of photons were large. In fact, the first successful theoretical description of these results was provided in 1978 by Leopold and Percival⁵⁶, who used the large quantum numbers in the experiment to justify a classical treatment of the Bohr atom in the oscillating electric field. By numerically solving the classical equations of motion they showed that the classical theory predicted ionization in good agreement with experiment.

The physical mechanism responsible for the classical ionization is the onset of chaos in this strongly perturbed nonlinear oscillator. In the absence of the perturbation, the classical equations of motion are integrable, and the bound electron motion is confined to a Kepler ellipse like the orbit of a planet about the Sun. But when the microwave perturbation is applied, the Kepler ellipse begins to wobble, precess and elongate, and when the perturbation exceeds a sharp threshold the electron escapes and the atom is ionized. For example, Fig. 2 illustrates this transition from regular, bounded motion to chaotic ionization in a one-dimensional classical model of a hydrogen atom in a strong oscillating field as the field is increased from below (Fig. 2a) to above (Fig. 2b) the threshold for widespread chaos.

Further confirmation²⁵ of this chaotic ionization mechanism was provided by experimental measurements of the threshold fields for the onset of microwave ionization over a wide range of excited states of hydrogen with principal quantum numbers from $n = 32$ to 90. Figure 6 shows a comparison of the experimental thresholds with the numerically calculated classical and quantum thresholds for the onset of chaotic ionization^{31,33,34} in a one-dimensional model of the experiment. The threshold microwave fields, F , for the measured onset of ionization are plotted in scaled atomic units, $n^4 F$, against the scaled microwave frequency, $n^3 \Omega$, for many different initial quantum states and two different microwave frequencies, 9.9 and 36 GHz. For both frequencies, the classical predictions for the onset of chaotic ionization are in remarkably good agreement with the quantum ionization thresholds (both experimental and theoretical) as long as the microwave frequency, Ω , is less than the Kepler orbital frequency, $1/n^3$ in atomic units. These results clearly indicate that, although quantum mechanics cannot satisfy the strict definitions of classical chaos, the quantum dynamics may nevertheless mimic the chaotic classical ionization, at least on the limited timescales of these experiments.

For higher frequencies, $n^3 \Omega > 1$, the classical and quantum thresholds start to diverge. This apparent breakdown of the correspondence principle was theoretically predicted by Casati *et al.*³⁷ who argued that the chaotic classical ionization should be suppressed by quantum interference effects related to the mechanism of Anderson localization in solid-state physics. As in the previous example of highly excited atoms in strong magnetic fields, these novel results for the interaction of intense

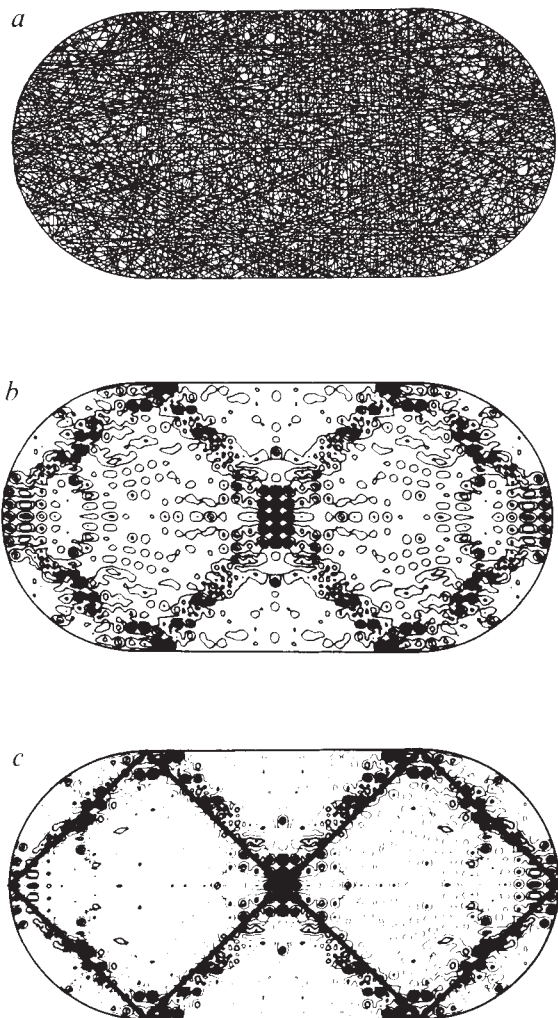


FIG. 5 a, Typical, chaotic classical trajectory for a 'stadium billiard'. Ten thousand bounces are shown of a highly irregular orbit, which never repeats and eventually will uniformly (ergodically) cover the area enclosed by the stadium walls. In stark contrast, highly excited quantum eigenfunctions often show regular structures that restrict the probability for finding the particle to bands along some of the unstable, periodic orbits in the stadium. A topographic map of one of these 'scarred' wavefunctions is shown in b, and the guiding unstable periodic orbit is superimposed in c. (Reprinted with permission from ref. 38.)

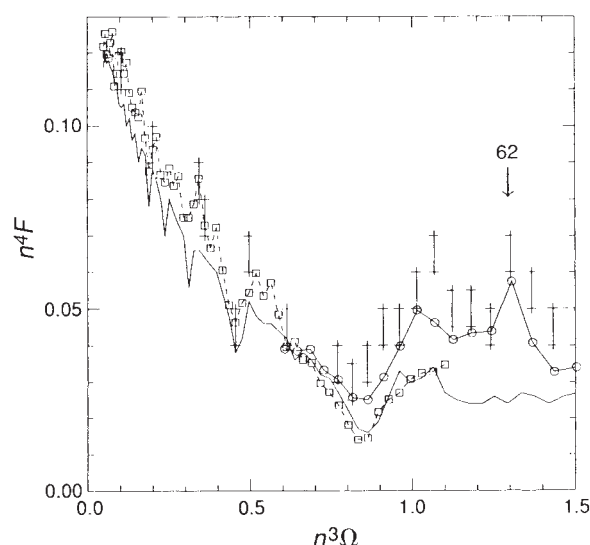


FIG. 6 Experimental measurements of the threshold microwave fields for the onset of ionization in scaled atomic units, $F_0 = n^4 F$, are plotted as a function of the scaled microwave frequency, $\Omega_0 = n^3 \Omega$, for highly excited hydrogen atoms with principal quantum numbers $n = 32$ to 90 , exposed to a 9.92-GHz field²⁵ ($\square$) and a 36 GHz field²⁶ ($\circ$). Each experimental data point indicates a different initial quantum state. For $\Omega_0 < 1$ these experimental results are in very good agreement with the classical predictions of the onset of chaotic ionization (solid curve) and the (less accurate) quantum calculations (crosses with large error bars) for a one-dimensional model of the experiment³³. For higher scaled frequencies, $\Omega_0 > 1$, the experimental thresholds continue to show good detailed agreement with the quantum calculations but at fields much larger than the classical thresholds. The arrow indicates the initial state, $n = 62$, that is stabilized by the excitation of the 'scarred' wavefunction shown in Fig. 7. (Reprinted with permission from ref. 60.)

electromagnetic fields with matter demand a careful reassessment of the correspondence principle.

In particular, for this time-dependent problem, the classical-quantum correspondence depends on a timescale known as the 'break-time', t_B which is roughly determined by the energy-level spacing of interacting quantum states, $\Delta E \propto 1/t_B$, through the Heisenberg uncertainty principle³⁵⁻³⁷. Roughly speaking⁵⁷, for short times $t < t_B$, the quantum system cannot recognize the discrete nature of the interacting quantum states with typical energy spacing ΔE , and the quantum evolution of observable quantities such as the expectation of position and momentum may be well approximated by classical dynamics. For longer times, the quantum discreteness of the coupled states requires that the quantum evolution of observables be quasiperiodic rather than chaotic.

For experiments at low scaled frequencies ($n^3 \Omega < 1$ in Fig. 6), many quantum states are strongly coupled by multiphoton processes of all orders, so the effective ΔE is small and the classical theory is a good approximation on the short timescales of the experiment. For higher frequencies, however, a dynamical selection rule restricts the coupling to relatively few quantum states that are close to resonance for single photon excitations^{58,59}. This reduction in the density of coupled states decreases the break-time so that quantum interference effects can inhibit the classical chaotic ionization on the timescales of the experiment.

High-scaled-frequency experiments^{24,26} have confirmed this prediction that the quantum system tends to be more stable against ionization than the classical. But the quantum excitation still seems to fluctuate considerably depending on the experimental parameters. For example, Fig. 6 shows that the quantum state with $n = 62$ is much harder to ionize than $n = 61$ or $n = 63$.

Theoretical work on the physical mechanisms responsible for this striking fluctuation in the quantum transport again illustrates the important role of scarred wavefunctions. Specifically, Jensen *et al.*⁶⁰ showed that this enhanced stabilization in the quantum calculations is due to the selective excitation of a quantum wavefunction that is highly localized to the vicinity of an unstable periodic orbit in the classical phase space. A topographic map of this scarred wavefunction is superimposed on the corresponding classical phase space in Fig. 7. The quantum probability for this state is strongly peaked near an unstable periodic orbit and its associated stable and unstable manifolds. In contrast, the wavefunctions excited from $n = 61$ and $n = 63$ are extended throughout the chaotic region of the classical phase space in Fig. 7 and therefore ionize more readily.

Applications and new directions

Quantum mechanics has traditionally (and very successfully) dealt with problems involving weakly interacting or weakly perturbed microscopic systems. Strongly coupled or strongly perturbed systems in atomic, molecular, solid-state, nuclear and elementary particle physics represent a new and difficult regime for quantum mechanics, requiring new theoretical as well as experimental methods.

In the past six years, theoretical and experimental investigations of hydrogen atoms in intense microwave fields or highly excited hydrogen and lithium atoms in strong magnetic fields have been the principal focus of much of the research in quantum chaos. Despite the relative simplicity of these physical systems, remarkable and unexpected properties of strongly perturbed quantum systems have appeared which have required a careful reassessment of our understanding of the classical-quantum correspondence principle. For example, the chaotic ionization of highly excited hydrogen atoms has shown how the quantum

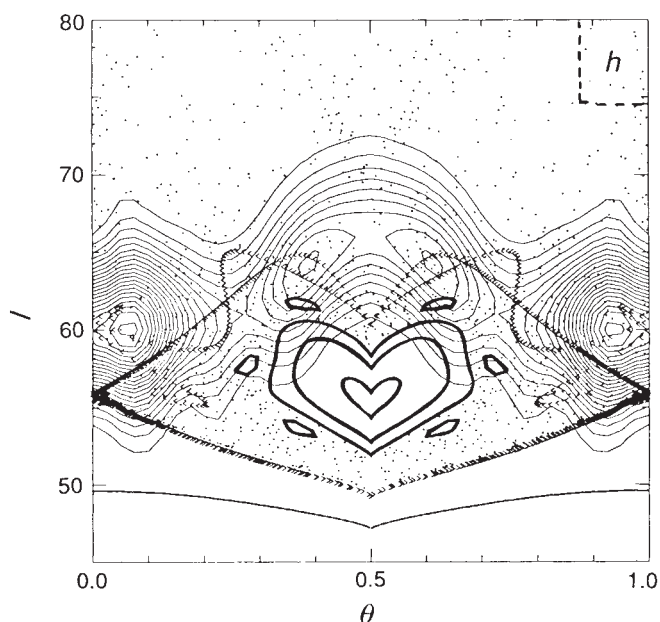


FIG. 7 A topographic map (light lines) of the highly localized, 'scarred' wavefunction that is responsible for the anomalous stability of $n = 62$ is superimposed on the corresponding classical Poincaré surface of section (dots and heavy lines) similar to those shown in Fig. 2. For the more experienced quantum chaologist, the relative size of Planck's constant for these experimental parameters is indicated in the upper right-hand corner, and the lines of arrows indicate the classical stable and unstable manifolds emanating from the unstable periodic orbit centred at $l = 56$ and $\theta = 0, 1$. (Reprinted with permission from ref. 60.)

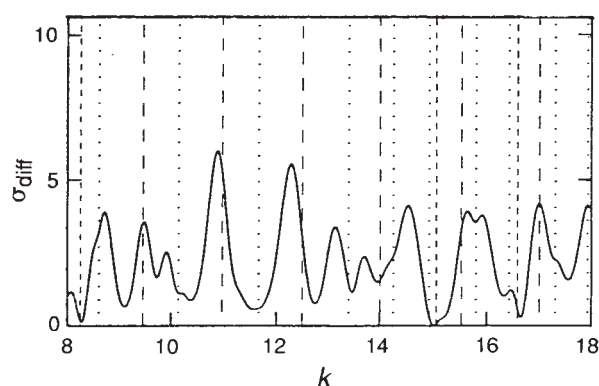


FIG. 8 The differential cross-section for the two-dimensional, elastic scattering of a plane wave from three perfectly reflecting discs, plotted as a function of incident wavenumber k . This scattering cross-section is only weakly correlated with the real parts of the resonance energies, indicated by the vertical broken lines, because the widths of the resonances are comparable to or greater than their separation. In this regime the irregular fluctuations of the scattering cross-section are reminiscent of the Ericson fluctuations in neutron scattering from heavy nuclei⁶³, the conductance fluctuations in small solid-state devices^{69,71} and the large fluctuations in the quantum transport measured by the threshold fields for microwave ionization for $n^3\Omega > 1$ in Fig. 6. (Reprinted with permission from ref. 67.)

dynamics can mimic the chaotic classical dynamics for limited times^{31,33,34} and the high-frequency stabilization of these Rydberg atoms has revealed the important role of quantum interference effects in localizing the quantum wavefunction^{35–37}, in close analogy with the mechanism of Anderson localization in solid-state physics. In addition, experimental measurements and theoretical calculations of the energy-level spectrum of highly excited atoms in strong, static magnetic fields have shown that the irregular spectrum for this simple quantum system with chaotic classical dynamics has the same statistical laws as the much more complicated and poorly understood spectrum of excited nuclei³². Finally, experiments on both physical systems have shown quantum wavefunctions that seem to have scars^{49,60} associated with unstable classical periodic orbits embedded in the chaotic classical phase space.

The lessons learned from these studies of single hydrogen atoms in strong fields promise to have many important applications in different areas of physics and chemistry. For example, in nuclear physics a realistic model of the low-lying collective states of nuclei⁶¹ (which is chaotic in a classical limit) has been found to have statistical distributions of energy-level spacings and transition amplitudes that are similar to those for the chaotic hydrogen atoms in strong magnetic fields. Moreover, in atomic physics it has been suggested⁶² that the quantum mechanical scarring of unstable periodic orbits may provide a mechanism for the suppression of ionization of ground-state atoms in super-intense laser fields.

Another exciting direction is the extension of these new ideas to irregular, or chaotic, scattering processes⁶³. Conventional quantum collision theory requires that the collisions be weak (so that perturbation theory can be applied) or that the collisions be very slow (adiabatic approximation) or fast (sudden approximation). If the collisions of electrons on atoms, atoms on molecules, molecules on surfaces, neutrons on nuclei or pions on protons are neither weak, nor slow, nor fast, then the scattering process may be very complex and difficult to describe. In the course of such a scattering event a collision complex is formed, the constituents rattle around, and the pieces fly apart. Chemists have already shown that the quantum description of this classically chaotic process can exhibit intricate interference

associated with the unstable periodic orbits in the collision complex, reminiscent of the scars in bounded atomic systems^{64–66}. Moreover, it has been shown that fluctuations in the collision cross-sections for electromagnetic waves in curved waveguides^{67,68} (shown in Fig. 8) and for electrons in mesoscopic solid-state devices⁶⁹ obey the same statistical laws found empirically for the scattering of neutrons from compound nuclei⁶³.

Like the oscillations of the photoabsorption spectrum of hydrogen atoms in strong magnetic fields and the fluctuations in the threshold fields for the microwave ionization of highly excited hydrogen atoms for $n^3\Omega > 1$, these so-called Ericson fluctuations⁶³ in the scattering cross-sections seem to be intimately related to the classical properties of unstable periodic orbits. Many experimental and theoretical studies of these interconnections between classical and quantum physics are in progress^{63–69}. For example, detailed comparisons of the classical, semiclassical and quantum theories with experimental measurements of the simplest inelastic and reactive atom-molecule collisions $D + H_2$ have recently been announced⁷⁰. In solid-state physics the efforts to fabricate smaller and smaller electronic devices have led to the discovery that electron scattering in 'mesoscopic' wires with widths smaller than $1\ \mu\text{m}$ produces 'universal conductance fluctuations', which are 'random' reproducible variations in the resistance or conductance of the wire⁷¹. These fluctuations even arise when the wires are so small that the electron collides only with the curved boundary of the wire⁶⁹. As a final illustration, it has also been suggested that the scattering cross-sections of elementary particles, such as pions on protons, may exhibit Ericson fluctuations⁷² due to the strong many-body interaction of the quark constituents. Since asymptotic freedom was discovered in the mid-1970s, most experimental emphasis has been on high-energy scattering, where the quarks are weakly coupled and perturbation theory can be applied⁷³, so this last extension of quantum chaos to the farthest boundaries of modern physics remains to be explored. $\square$

Roderick Jensen is in the Department of Physics, Wesleyan University, Middletown, Connecticut 06457, USA.

- May, R. M. *Nature* **261**, 459–467 (1976).
- Dynamical Chaos* (eds Berry, M. V., Percival, I. C. & Weiss, N. O.) (The Royal Society, London, 1987).
- The Ubiquity of Chaos* (ed. Krasner, S.) (American Association for the Advancement of Science, Washington DC, 1990).
- Jensen, R. V. *Am. Scient.* **75**, 168–181 (March 1987); *Encyclopedia of Modern Physics*, 69–96 (Academic, New York, 1990).
- Ford, J. in *The New Physics* (ed. Davies, P. C. W.) 348–371 (Cambridge University Press, 1986).
- Gleick, J. *Chaos: Making a New Science* (Viking, New York, 1987).
- Goldstein, H. *Classical Mechanics* (Addison-Wesley, Reading, Massachusetts, 1980).
- Lichtenberg, A. J. & Leiberman, M. A. *Regular and Stochastic Motion* (Springer, New York, 1983).
- Wisdom, J. *Nature* **315**, 731–733 (1985).
- Poincaré, H. *Les Methodes Nouvelles de la Mécanique Celeste* (Gauthier-Villars, Paris, 1892); (reprinted) vol. 3, section 297 (Dover, New York, 1957).
- Delos, J. B., Knudson, S. K. & Noid, D. W. *Phys. Rev. A* **28**, 7–21 (1983).
- Bohr, N. *Kgl. Danske Vid. Selsk. Skr., nat-math. Afd. B. Raekke* IV.1, 1–3 (1918); (reprinted in *Sources of Quantum Mechanics* (ed. van der Waerden, B. L.) 95–137 (Dover, New York, 1967).
- Berry, M. V. *Proc. R. Soc. Lond. A* **412**, 183–198 (1987).
- Chirikov, B. V., Israiliev, F. M. & Shepelyansky, D. L. *Physica D* **33**, 77–88 (1988).
- Chaos and Quantum Physics* (eds Giannoni, M.-J., Voros, A. & Zinn-Justin, J.) (Elsevier, London, 1991).
- Holle, A., Wiebush, G., Main, J., Friedrich, H. & Welge, K. H. *Phys. Rev. Lett.* **56**, 2594–2597 (1986).
- Holle, A. *et al.* *Z. Phys. D* **5**, 279–285 (1987).
- Main, J., Wiebush, G., Holle, A. & Welge, K. H. *Phys. Rev. Lett.* **57**, 2789–2792 (1986).
- Holle, A., Main, J., Wiebush, G., Rottke, H. & Welge, K. H. *Phys. Rev. Lett.* **61**, 161–164 (1988).
- Welch, G. R., Kash, M. M., Iu, C., Hsu, L. & Kleppner, D. *Phys. Rev. Lett.* **62**, 893–896 (1989).
- Iu, C., Welch, G. R., Kash, M. M., Hsu, L. & Kleppner, D. *Phys. Rev. Lett.* **63**, 1133–1136 (1989).
- Iu, C., Welch, G. R., Kash, M. M., Kleppner, D., Delande, D. & Gay, J.-C. *Phys. Rev. Lett.* **66**, 145–148 (1991).
- Bayfield, J. E. & Koch, P. M. *Phys. Rev. Lett.* **33**, 258–261 (1974).
- Bayfield, J. E., Casati, G., Guarneri, I. & Sokol, D. W. *Phys. Rev. Lett.* **63**, 364–367 (1989).
- van Leeuwen, K. A. H. *et al.* *Phys. Rev. Lett.* **55**, 2231–2234 (1985).
- Galvez, E. J., Sauer, B. E., Moorman, L., Koch, P. M. & Richards, D. *Phys. Rev. Lett.* **61**, 2011–2014 (1988).
- Koch, P. M. in *The Ubiquity of Chaos* (ed. Krasner, S.) 75–97 (American Association for the Advancement of Science, Washington DC, 1990).
- Friedrich, H. & Wintgen, D. *Phys. Rep.* **183**, 37–79 (1989).
- Hasegawa, H., Robnik, M. & Wunner, G. *Progr. theor. Phys.* **98**, 198–286 (1989).
- Delande, D. in *Chaos and Quantum Physics* (eds Giannoni, M.-J., Voros, A. & Zinn-Justin, J.) (Elsevier, London, 1991).
- Jensen, R. V. *Phys. Rev. A* **30**, 386–397 (1984).
- Quantum Chaos and Statistical Nuclear Physics* (eds Seligman, T. H. & Nishioka, H.) (Springer, New York, 1986).

33. Jensen, R. V., Susskind, S. M. & Sanders, M. M. *Phys. Rep.* **201**, 1–56 (1991).
34. Blümel, R. & Smilansky, U. *Z. Phys.* **D6**, 83–105 (1987).
35. Casati, G., Chirikov, B. V. & Shepelyansky, D. L. *Phys. Rev. Lett.* **53**, 2525–2528 (1984).
36. Casati, G., Guarneri, I. & Shepelyansky, D. L. *IEEE J. Quant. Electron.* **24**, 1420–1444 (1988).
37. Chirikov, B. V. in *Chaos and Quantum Physics* (eds Giannoni, M.-J., Voros, A. & Zinn-Justin, J.) (Elsevier, London, 1991).
38. Heller, E. J. *Phys. Rev. Lett.* **53**, 1515–1518 (1984).
39. Zeeman, P. *Phil. Mag. Ser. 5* **43**, 226–239 (1897).
40. Schiff, L. I. *Quantum Mechanics*, 283–288 (McGraw-Hill, New York, 1955).
41. Garstang, R. H. *Rep. Prog. Phys.* **40**, 105–154 (1977).
42. Percival, I. C. *Adv. chem. Phys.* **36**, 1–62 (1977).
43. Zaslavsky, G. M. *Phys. Rep.* **80**, 157–250 (1981).
44. Einstein, A. *Ver. Deut. Phys. Ges.* **19**, 82–92 (1917).
45. Garton, W. R. S. & Tompkins, F. S. *Astrophys. J.* **158**, 839–845 (1969).
46. Du, M. L. & Delos, J. B. *Phys. Rev. A* **38**, 1896–1912 (1988).
47. Macdonald, J. & Kaufman, A. *Phys. Rev. Lett.* **42**, 1189–1191 (1979).
48. Waterland, R. L., Yuan, J.-M., Martens, C. C., Gillilan, R. E. & Reinhardt, W. P. *Phys. Rev. Lett.* **61**, 2733–2736 (1988).
49. Wintgen, D. & Hönig, A. *Phys. Rev. Lett.* **63**, 1467–1470 (1989).
50. Sirdhar, S. *Phys. Rev. Lett.* **67**, 785–788 (1991).
51. Gutzwiller, M. *Chaos In Classical and Quantum Mechanics* (Springer, New York, 1990).
52. Feynman, R. P. & Hibbs, A. R. *Quantum Mechanics and Path-Integrals* (McGraw-Hill, New York, 1965).
53. Ezra, G. S., Richter, K., Tanner, G. & Wintgen, D. *J. Phys. B* **24**, L413–L420 (1991).
54. Van Vleck, J. H. *Phil. Mag.* **44**, 842–869 (1922).
55. Einstein, A. *Ann. Phys.* **17**, 132–148 (1905).
56. Leopold, J. G. & Percival, I. C. *Phys. Rev. Lett.* **41**, 944–947 (1978).
57. Chirikov, B. V., Israilev, F. M. & Shepelyansky, D. L. *Sov. Scient. Rev. C2*, 209 (1981).
58. Jensen, R. V., Susskind, S. M. & Sanders, M. M. *Phys. Rev. Lett.* **62**, 1476–1479 (1989).
59. Leopold, J. G. & Richards, D. J. *Phys. B22*, 1931–1961 (1989).
60. Jensen, R. V., Sanders, M. M., Saraceno, M. & Sundaram, B. *Phys. Rev. Lett.* **63**, 2771–2775 (1989).
61. Alhassid, Y., Novoselsky, A. & Whelan, N. *Phys. Rev. Lett.* **65**, 2971–2974 (1990).
62. Jensen, R. V. & Sundaram, B. *Phys. Rev. Lett.* **65**, 1964 (1990).
63. Smilansky, U. in *Chaos and Quantum Physics* (eds Giannoni, M.-J., Voros, A. & Zinn-Justin, J.) (Elsevier Science, London, 1991).
64. Pollak, E. & Child, M. S. *Chem. Phys.* **60**, 23 (1981).
65. Thompson, T. C. & Truhlar, D. G. *J. phys. Chem.* **88**, 210–214 (1984).
66. Davis, M. J. *J. phys. Chem.* **92**, 3124–3144 (1988).
67. Gaspard, P. & Rice, S. *J. chem. Phys.* **90**, 2225–2241; 2242–2254; 2255–2262 (1989).
68. Doron, E., Smilansky, U. & Frenkel, A. *Phys. Rev. Lett.* **65**, 3072–3075 (1990).
69. Jalabert, R. A., Baranger, H. U. & Stone, A. D. *Phys. Rev. Lett.* **65**, 2442–2445 (1990).
70. Kliner, D. A. V., Rinnen, K.-D. & Zare, R. N. *Chem. Phys. Lett.* **166**, 107–111 (1990).
71. Lee, P. A., Stone, A. D. & Fukuyama, H. *Phys. Rev. B* **35**, 1039–1070 (1987).
72. Frautschi, S. C. *Nuovo Cimento A* **12**, 133–161 (1972).
73. Politzer, H. D. *Phys. Rep.* **14**, 129–180 (1974).
74. Delande, D. thesis, Univ. Pierre and Marie Curie, Paris (1988).

ACKNOWLEDGEMENTS. I thank my experimental and theoretical colleagues, N. Balasz, J. Bayfield, M. Berry, R. Blümel, G. Casati, B. Chirikov, J. Delos, S. Frautschi, I. Guarneri, S. Fishman, J. Ford, M. Gutzwiller, E. Heller, D. Kleppner, P. Koch, J. Leopold, D. Richards, I. Percival, R. Prange, M. Saraceno, B. Sundaram, U. Smilansky, K. Welge and D. Wintgen for stimulating discussions. This work was supported by the NSF.

ARTICLES

The structure of the *E. coli* recA protein monomer and polymer

Randall M. Story, Irene T. Weber* & Thomas A. Steitz

Department of Molecular Biophysics and Biochemistry, and Department of Chemistry and Howard Hughes Medical Institute at Yale University, New Haven, Connecticut 06511, USA

The crystal structure of the *recA* protein from *Escherichia coli* at 2.3-Å resolution reveals a major domain that binds ADP and probably single- and double-stranded DNA. Two smaller subdomains at the N and C termini protrude from the protein and respectively stabilize a 6₁ helical polymer of protein subunits and interpolymer bundles. This polymer structure closely resembles that of *recA*/DNA filaments determined by electron microscopy. Mutations in *recA* protein that enhance coprotease, DNA-binding and/or strand-exchange activity can be explained if the interpolymer interactions in the crystal reflect a regulatory mechanism *in vivo*.

THE *recA* protein (*recA*) has a pivotal role in homologous DNA recombination and DNA repair in *E. coli* and other eubacteria. The gene encoding *recA* was first shown in 1965 to be essential for genetic recombination and resistance to ultraviolet irradiation¹. The 352-residue protein^{2,3} encoded by this gene has a remarkable range of activities *in vitro* (for reviews, see refs 4–6). *RecA* protein alone catalyses the central steps of recombination: the pairing and strand exchange of homologous DNA molecules.

RecA protein polymerizes cooperatively and nonspecifically on DNA to form a helical filament that is the active species in the DNA-strand exchange and repressor-cleavage reactions. Filaments are formed by rapid polymerization 5' to 3' on single-stranded DNA or duplex DNA possessing a single-stranded

gap⁷. Stable binding to duplex DNA is also observed, but this process is slow at neutral pH (ref. 8). Formation of active *recA* filaments requires ATP or an analogue such as ATP-γ-S⁹. Such filaments can then specifically bind homologous duplex DNA and promote a strand-exchange reaction¹⁰. They can also induce the cleavage of *lexA* (which leads to the SOS response¹¹), *umuD*¹² and the repressor proteins of λ and other phages, binding these proteins and stimulating their intrinsic ability to self-cleave¹³, thereby acting as a 'coprotease'.

Image reconstructions of electron micrographs have provided a low-resolution structure of the *recA* protein filament bound to single- or double-stranded DNA^{14,15}. These filaments have a pitch of 85–100 Å, 6.1–6.2 monomers per turn, and a distinct polarity. The DNA in *recA*/DNA filaments is extended by ~150% relative to B-form DNA, and unwound to ~18.6 nucleotides (or base pairs) per turn¹⁶. The stoichiometry of binding is three nucleotides (or base pairs) per *recA* monomer¹⁷. Geometrical constraints require that the DNA must lie near the centre of the helical filament¹⁴. In the absence of DNA, *recA* protein exists in various states of aggregation, including pure protein polymers¹⁸. These polymers have a shorter pitch than the active *recA*/DNA filaments (~70 Å; refs 19, 20), but under certain conditions, such as high salt, they assume the higher pitch and have the ATPase and coprotease activities normally associated with the *recA*/DNA filament^{21,22}.

For a full understanding of *recA* protein function we need details of the structure of the *recA* filament because of its importance in *recA* activity. We have established and refined at 2.3 Å resolution the crystal structure of the *recA* protein that forms a *recA* polymer in the crystal. This polymer closely resembles the low-resolution structure of *recA*/DNA filaments seen in the electron microscope, and contains an extensive interface between subunits. The study of *recA* structure and function has been complicated by the interrelatedness of the

* Present address: Department of Pharmacology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA.

scattering difference Fourier map calculated with native structure factors and phases derived from the model, the positions of the eight highest peaks correspond to eight of the nine sulphurs in the model (Fig. 1). This map establishes the correctness of the polypeptide chain trace for at least the first 175 residues.

Description of the structure

The structure of recA protein consists of a major central domain flanked by two smaller subdomains at the amino and carboxy termini (Fig. 2*a, b*). The major domain of the molecule contains a mostly parallel, twisted, 8-stranded β -sheet that is flanked by α -helices. As in many other mono- and dinucleotide-binding proteins, the nucleotide binds near the carboxy end of the mostly parallel β -sheet, with its phosphates bound near the amino-terminus of an α -helix²⁵ (Fig. 2, and ref. 26). As no electron density corresponding to residues 157–164 and 195–209 (loops L1 and L2) is visible in the electron density map, they are presumably disordered. As described later, these regions lie towards the filament axis and are proposed to be involved in DNA binding. A turn between two of the antiparallel β -strands, residues 232–236, also lies in an area of weak electron density and is ill-defined.

Two smaller subdomains at the amino and carboxy ends of the molecule protrude from the protein and stabilize formation of the polymer and of interpolymer bundles respectively. The amino-terminal region of the molecule, from residues 1 to about 30, forms α -helix A and β -strand 0 and protrudes from the rest of the molecule. This α/β unit is important in forming the recA polymer (see below). The small C-terminal domain, which includes residues from about 270 to 328, consists of three helices and two β -strands. This domain also protrudes from both the protein and from the recA protein polymer, making stabilizing interactions between polymers in the crystal (see below). The final 25 residues are not visible in the electron density map and are presumably disordered.

The structure of the nucleotide-binding core of recA protein was predicted²⁷, assuming a six-stranded parallel sheet with classic 'Rossmann fold' topology²⁸. These assumptions have proved to be false, and the model incorrect. The structure also disagrees with proposals that recA protein consists of two approximately equally sized domains^{15,29}.

Polymer structure

RecA protein packs along the 6₁ axis in the P6₁ crystals to form a polymer that is ~120 Å wide, with a pitch of 82.7 Å, which corresponds to the *c* cell dimension (Fig. 3*a, b*). Monomers are packed together so as to form a continuous spiral of protein, with the globular carboxy terminal domain projecting outwards from the rest of the polymer. The carboxy end of the parallel β -sheet points inwards towards the centre of the polymer so that regions around disordered loops L1 and L2 (residues 157–164 and 195–209) lie closest to the polymer axis. The position of ADP binding²⁶ is near the inner surface of the polymer, oriented so that the phosphates point inwards. The helical polymer is rather open, with no contacts between successive turns. This forms a large helical groove, which is considerably larger than the hole of ~25 Å diameter seen in projection looking down the 6₁ axis. One surface bordering the helical groove (the upper surface in Fig. 3*a*) is more or less 'smooth', formed by helices A, G, F and E, whereas the other has a notched appearance formed by lobes that hang down and which are composed in part of the carboxy terminal domain.

The interface between subunits in the polymer is very extensive. Part of this interaction is formed by packing of β -strand 0 and α -helix A, near the amino terminus of the protein molecule, in an antiparallel fashion against β -strand 3 and α -helix E, respectively, of an adjacent molecule. Such packing effectively adds another α/β unit to the α/β domain of an adjacent molecule, extending the β -sheet to nine strands. This

TABLE 1 Heavy-atom refinement statistics

Derivative	Resolution (Å)	R_{sym} (%)	R_F (%)	Refined sites	Site 1:site 2*	f_H/E
Native	2.7	4.74	—	—	—	—
Pt(NO ₂) ₄	3.0	5.96	13.0	6	2.33	1.81
PtCl ₄	3.0	7.38	15.6	4	1.22	1.84
PtCl ₆	3.0	8.73	12.7	4	0.96	1.80
PtBr ₄	3.0	10.03	15.3	4	0.80	1.41
Hg(CH ₃) ₂	3.5	8.53	12.7	4	—	0.70

Mean figure of merit versus resolution

>11.7 Å 8.28 Å 6.40 Å 5.22 Å 4.40 Å 3.81 Å 3.36 Å 3.00 Å overall
0.85 0.91 0.87 0.78 0.70 0.59 0.49 0.36 0.58

$$R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I, R_F = \sum |F_{\text{deriv}} - F_{\text{native}}| / \sum F_{\text{native}}$$

$$f_H/E = [\sum f_H^2 / \sum (F_{\text{deriv,obs}} - F_{\text{deriv,calc}})^2]^{1/2}$$

RecA protein is overexpressed by inducing with nalidixic acid cells containing the recA⁺ plasmid pDR1453 (ref. 62). Following cell lysis, protein is purified by polymin P-precipitation, ammonium sulphate precipitation and hydroxylapatite chromatography. Purified protein is resuspended at ~20 mg ml⁻¹ in a solution containing 0.1 M citrate-phosphate buffer, pH 5.4, 0.1 M NaCl, 10% glycerol, 0.1 mM dithiothreitol, 0.02% Na₂S₂O₃ and 0.5 or 1% polyethylene glycol (PEG 8000). Crystals grow in several days by dialysis versus 2 or 3% PEG 8000 at 25°C. For data collection, crystals were transferred stepwise to 0.1 M MES or acetate, pH 5.4, 50% ethylene glycol, 0.05 M NaCl and 0.02% Na₂S₂O₃. Platinum soaks were all at 0.1 mM for 1–2 days. All data were collected from single crystals on a Xuong-Hamlin area detector and processed with software from the University of California at San Diego. Initial phases were determined by refining the Pt(NO₂)₄ derivative alone using the program PHARE (of the CCP4 package, Daresbury, UK). These phases were modified by combining them with those from a backtransformed solvent-flattened electron density map³³ (assuming a conservative solvent content of 52%). They were then used as fixed external phases in the refinement of the heavy-atom parameters. Each additional derivative was added individually and refined in this way until it converged (generally about 5 solvent-flattening/refinement cycles). Finally, minor sites were added and the heavy-atom temperature factors of major sites refined anisotropically. The final refined heavy-atom parameters were used to calculate a solvent-flattened electron density map to 3 Å resolution. Although the electron density map at this stage was of sufficient quality to allow identification of nearly all the secondary structure elements, we were unable to fit the sequence owing to the presence of disordered regions, the extensive contact between adjacent molecules in the polymer, and poorly defined side-chain electron density. Therefore, we used model building to improve the phases and extend them to 2.7 Å resolution. An initial fit of 290 residues of polyalanine gave an *R*-factor of 48.4% from 20–2.7 Å resolution. Calculation of electron density maps with phases derived from this model or from combined model and experimental phases showed a significant improvement in the side-chain and main-chain electron density. A subsequent rebuilding of the model and fitting of some of the residues with side chains of the appropriate size gave an *R*-factor of 44.8%. This model was refined using X-PLOR³⁴ positional refinement and rebuilt and refined once more to give an *R*-factor of 33.2%. At this stage we were able to fit the sequence unambiguously into the electron density. Refinement of the model then proceeded for several cycles using X-PLOR (including simulated annealing) and manual rebuilding, initially at 2.7 Å resolution and later at 2.3 Å resolution. The present model is refined against native data collected on 6 crystals, with $R_{\text{sym}} = 5.5\%$, 92% complete to 2.3 Å resolution.

* Relative occupancies of the two major platinum sites.

interaction, however, accounts for only part of the polymer interface, which is formed by two complementary faces of the molecule. The total loss in surface area accessible to solvent³⁰ in going from monomer to polymer is 2,890 Å² per subunit, assuming the structure of an isolated monomer is unchanged from that found in the polymer. This is in the range previously observed in oligomeric proteins³¹ and accounts for 19.4% of the total accessible surface area of a monomer, excluding the disordered regions that are not included in the model.

The 55 residues whose solvent-accessible surface area would decrease by >15 Å² on polymer formation are shown in Fig.

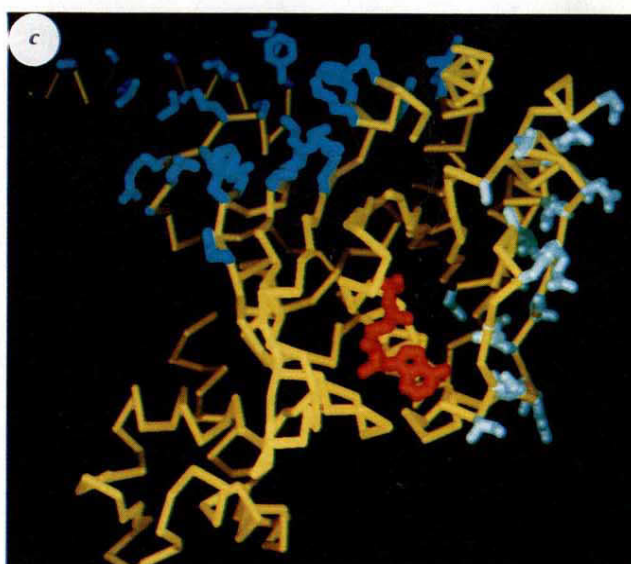
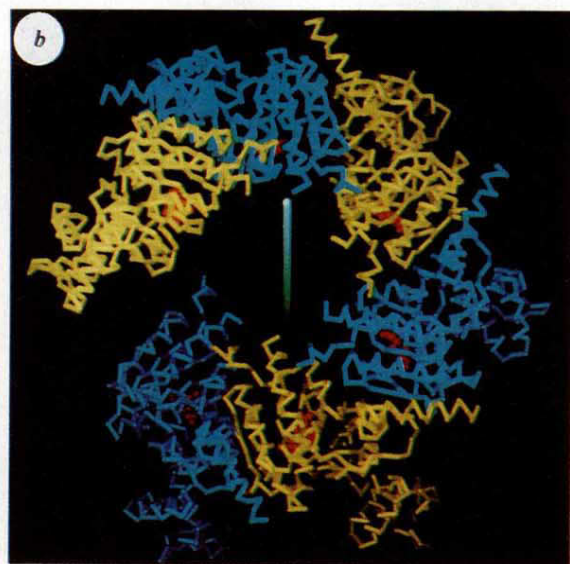
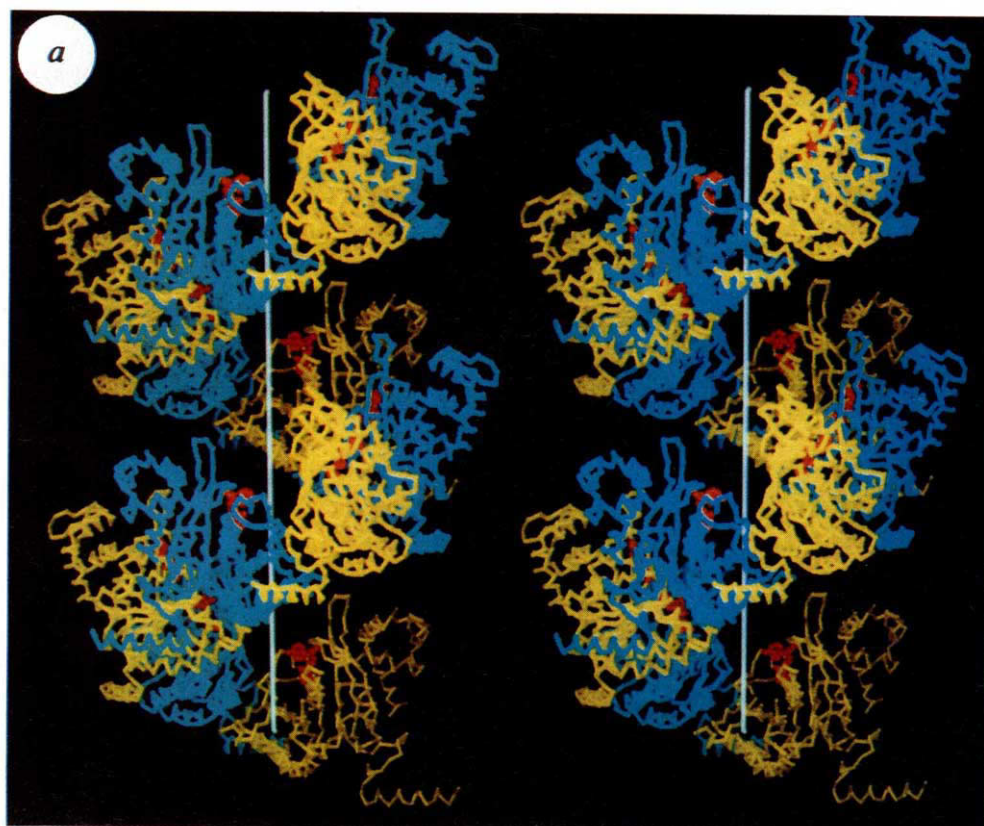
3c. The diverse interactions include a variety of hydrophobic and polar interactions. In fact, of the 20 amino acid residues, all but tryptophan are found at least once in the interface. Also, there are several well-ordered water molecules largely buried in the interface, connected by hydrogen-bonded networks to the bulk solvent.

Comparison with recA/DNA filament

A comparison between the recA polymer found in the crystal and the low-resolution electron microscopy image of the

recA/duplex DNA filament^{14,15} reveals a striking similarity. Like the polymer described here, the nucleoprotein filament has a large helical groove which is bordered on one side by a 'smooth' face and on the other by lobes that project downwards and outwards. The pitch is somewhat larger, roughly 85–100 Å, as is the number of subunits per turn, approximately 6.1–6.2. It is likely that the allosteric state of the recA protein polymer in the crystal is different from that of the active (ATP-bound) recA/DNA filament. But the large, highly complementary subunit interface observed in the crystal strongly suggests that

FIG. 3 The helical recA polymer observed in the crystal. *a*, Stereo α -carbon representation viewed parallel to the crystallographic 6_1 screw axis (shown in white). 12 monomers (two turns) are shown in alternating blue and yellow, with the bound ADP in red. The orientation of the upper left monomer is similar to that in Fig. 2. *b*, View looking (nearly) down the 6_1 axis; one turn is shown. The chain termini visible near the centre are the edges of the disordered region L1 (residues 157–164). The protruding carboxy terminal domain can be seen in the lower part of the figure. *c*, Residues involved in forming the two surfaces of the interface. Because of the asymmetric nature of interactions between subunits of a polymer, each subunit has two separate surfaces that form the contact interface. Those 55 residues whose solvent-accessible surface area decreases by $15 \text{ \AA}^2$ or more upon polymer formation (using a water radius of $1.4 \text{ \AA}$) are shown. Residues meeting this criterion are 6, 9, 10, 13, 14, 17, 20, 21, 25–30, 172, 176, 213, 216–218, 222, 247, 248, 250, 254, 255 and 257 on one face (in blue), and 89, 95, 97, 99, 101, 102, 111–120, 124, 127, 128, 131, 132, 135, 137, 139, 148, 150, 155 and 156 on the other face (in white). Side chains are shown even where main-chain atoms mostly or wholly contribute to the interface. 63% of the interface is made up of hydrophobic (carbon) atoms. The orientation of this view is the same as in Fig. 2.



although the form of the polymer we find in the crystal may differ from the recA/ATP/DNA filament in some details, the general features of its structure will be quite similar.

The polarity of the recA/DNA filament relative to single-stranded DNA is known³². As we can clearly align the polarity of the crystal polymer with the electron microscopy images of the recA/DNA filament, we deduce that single-stranded DNA would run 5' to 3' from top to bottom relative to the orientation shown in Fig. 3a. The direction of filament assembly (and also strand exchange) is 5' to 3' relative to DNA⁷, thus growth would

occur at the bottom of the polymer in this view. This observation, combined with what we know about the interface, explains the results of Yarranton and Sedgewick³³, who showed that a plasmid containing the amino terminal 22% (~78 residues) of recA protein negatively complements recA⁺ cells. Such a fragment, containing α -helix A and β -strand 0, presumably poisons filament growth by adding to the growing end and preventing further polymerization.

DNA binding sites

A key component in understanding the mechanism of recA action is the location of the site(s) of DNA binding. RecA protein is thought to have two DNA binding sites that can bind two DNA molecules even in the absence of complementarity^{34,35}—one for primary binding (to single-stranded or gapped duplex DNA) to form recA/DNA filaments, the other for secondary binding to homologous duplex DNA. We expect that regions involved in DNA binding are on the surface of the recA filament and highly conserved. A recent compilation of 16 bacterial recA proteins shows that 105 residues (~30%) are invariant⁴. Strikingly, invariant residues cluster on the inside of the recA polymer found in the crystal (Fig. 4a), which is not surprising because DNA must lie near the centre of the helical recA/DNA filament¹⁴. The disordered loops that lie near the polymer axis are also highly conserved, with 10 of 23, or 44% invariant residues.

Several lines of evidence lead us to conclude that residues in or around the disordered regions L1 (157–164) and L2 (195–209) form the binding site(s) for DNA (Fig. 4b). These regions lie closest to the polymer axis, and thus the site(s) of DNA binding. Loop L2 projects out from the carboxy end of the β -sheet, and lies towards the innermost face of the polymer directly above the ATP binding site. The recA430 mutation (Gly 204 $\rightarrow$ Ser)³⁶ leads to defects in repressor cleavage and ultraviolet resistance *in vivo*³⁷. *In vitro*, this mutant protein has a lower affinity for single-stranded DNA than wild type and is inhibited by single-stranded DNA-binding protein³⁸. Other mutations in this region (Glu 207 $\rightarrow$ Gln and Gly 211 $\rightarrow$ Ala) completely eliminate recA activity *in vivo*³⁹, consistent with its being involved in binding

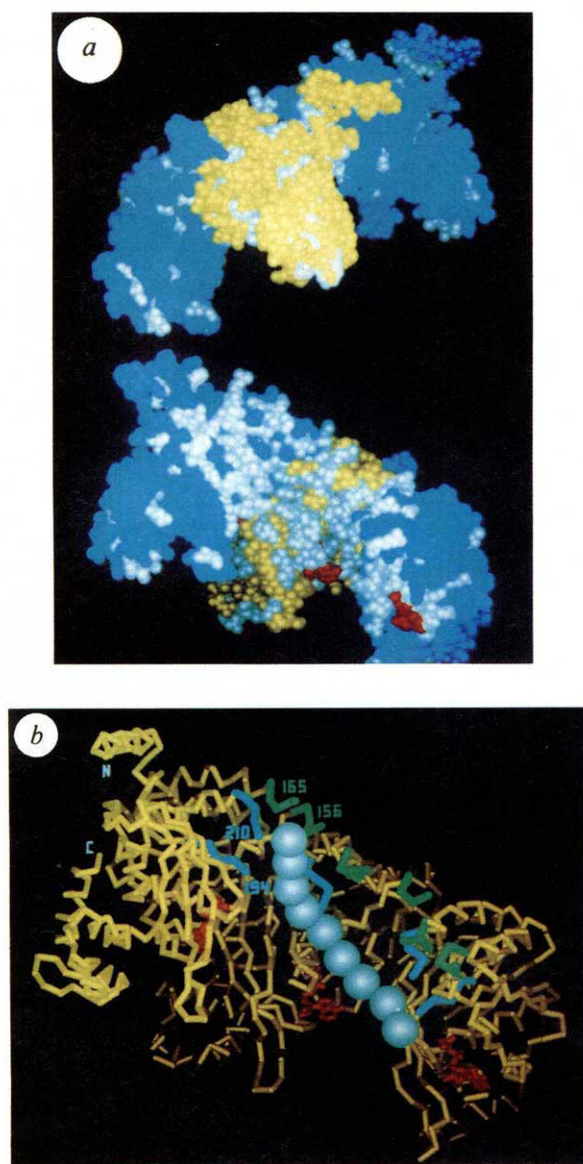


FIG. 4 Region that may be involved in DNA binding. *a*, Space-filling representation of the residues that are absolutely invariant among 16 recA sequences. Three subunits (one-half turn of the polymer) are shown viewed from the inside (lower) or outside (upper) of the polymer. Alternating monomers are coloured blue and yellow. Invariant residues are shown as white, ADP is red. The 6₁ axis is vertical in this view. As the exterior of the filament contains few invariant residues, models for recA function that postulate secondary DNA binding to the outside of the filament are unlikely to be correct. *b*, Position of the disordered loops L1 (156–165) and L2 (194–210). 5 α -carbons preceding and following each disordered loop are coloured green (L1) and blue (L2). α -carbons of 3 monomers (one-half turn of the polymer) are shown, with the 6₁ axis vertical. ADP is shown in red. A plausible model for the position of the single-stranded DNA phosphate backbone is depicted as white spheres of radius 3.5 Å located at a radius of 11.5 Å from the 6₁ axis and a stoichiometry of 3 nucleotides per recA monomer.

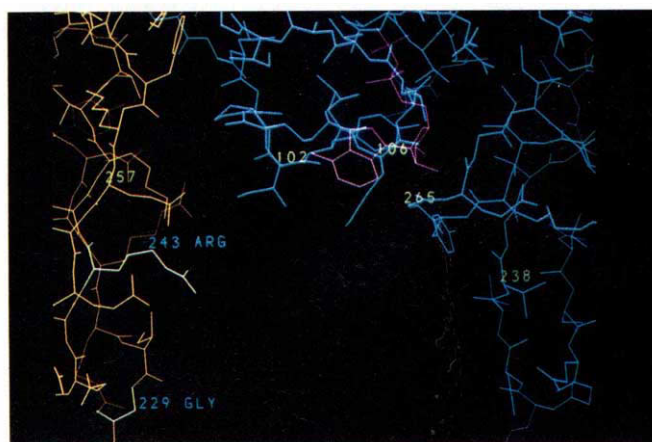


FIG. 5 Possible binding site for lexA, umuD and phage repressors. Atoms from two adjacent molecules are shown in yellow and blue; ADP is purple. Residues 229 and 243 whose mutation affects repressor cleavage are labelled. The 6₁ screw axis is vertical in this view. Directly above this region is the interaction between β -strand 0 and β -strand 3 of adjacent monomers in the polymer. Mutation of residues in this interface affects repressor cleavage *in vivo*. The recA1222 (Ser 25 $\rightarrow$ Phe)⁵³ and Cys 116 $\rightarrow$ Ser⁶⁶ mutations yield coprotease constitutive phenotypes, whereas the recA1730 mutation (Ser 117 $\rightarrow$ Phe) results in a phenotype deficient in lexA, but not ϕ_{80} cleavage⁴⁸. The coprotease constitutive mutant proteins might perturb polymerization so as to affect the equilibria between aggregates of protein and recA/DNA filaments (step 2 in Fig. 6c). The explanation for the phenotype of the recA1730 mutant is unclear.

single-stranded DNA, although there are other possible explanations.

Helix G is an attractive region to implicate in single-stranded DNA binding as it is centred on the most conserved region of sequence of bacterial *recA* proteins⁴. At the amino terminus of this α -helix are two glycines (211 and 212), which could permit favourable interactions between DNA phosphates and the helix dipole and backbone NH groups, as in many mono- and dinucleotide phosphate-binding sites⁴⁰. These two glycines are absolutely invariant in all known bacterial *recA* proteins⁴ and other *recA* homologues: T4 *UvsX* (ref. 4), and yeast *dmc1* (D. Bishop and N. Kleckner, personal communication). Alternatively, these two conserved glycines could be playing a structural part in mediating the ATP-induced conformational change.

Residues around disordered loop L1 (residues 156–165) form the inner, upper lip of the polymer as viewed in Fig. 3a and may form the binding site for homologous duplex DNA. The residues in this vicinity form one face of the helical groove likely to be accessible in *recA*/DNA filaments. Loop L1 is also the site of residues whose mutation affects DNA binding. The *recA1* mutation (Gly 160→Asp)³⁶ is completely inactive *in vivo*¹, and *in vitro* at neutral pH binds single-stranded DNA weakly and is unable to hydrolyse ATP⁴¹. Unlike wild-type protein, the *recA1* mutant protein does not bind a second strand of DNA in the presence of ATP- γ -S⁴². Studies on a Asn-160 mutation⁴³ and a nearby His 163→Ala change⁴⁴ indicate that such mutations prevent an ATP-induced conformational change necessary for strand exchange. Taken together, these data indicate that residues in or around loop L1 may be involved in the secondary binding of homologous duplex DNA necessary for recombination. This conclusion is also supported by two other mutations

in this region, *recA1602* (Gly 157→Asp) and *recA1203* (Arg 169→Cys), which give coprotease constitutive phenotypes, with defects in recombination⁴⁵. Although the reason for the coprotease constitutive phenotype is unclear, it does suggest that binding to single-stranded DNA is as good or better than wild type, with the recombination deficiency being due to a defect later in the strand-exchange reaction.

The structure of the polymer does not support suggestions that residues near the amino terminus⁴⁶ or residues 243–257 (ref. 47) are involved in DNA binding. Both of these regions are distant from the polymer axis and are involved in packing interactions between monomers in the polymer.

Several factors complicate model building of DNA into the *recA* polymer. The structure of the highly unwound and extended DNA molecule is not known. As already explained, a portion of the protein region likely to contact the DNA is not visible in the electron density map, so complete protein–DNA contacts cannot be predicted. Finally, some aspects of the ATP-filament structure pertinent to DNA binding must differ from the structure seen here. Although such extended, unwound DNA can be arranged in the polymer in several ways (for example, Fig. 4b), the ambiguities still preclude a meaningful molecular model for DNA binding.

Repressor binding site

The location of amino-acid substitutions in *recA* mutant proteins that are altered in their ability to function as a coprotease in *lexA* and repressor cleavage suggests that the repressor binding site may be in the 'notch' region between adjacent lobes of the *recA* polymer (Fig. 5). Two mutations that differentially affect cleavage of repressors—*recA1734* (Arg 243→Leu)⁴⁸ and *recA91*

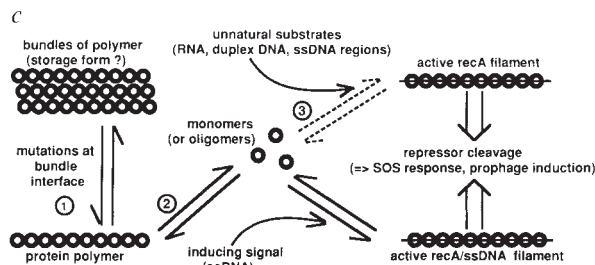
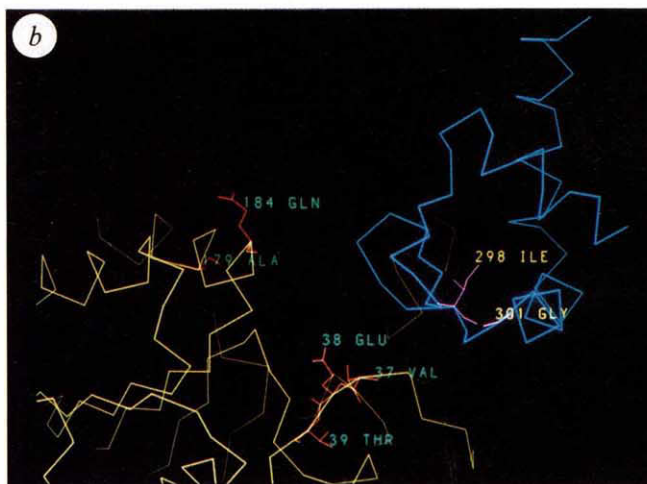
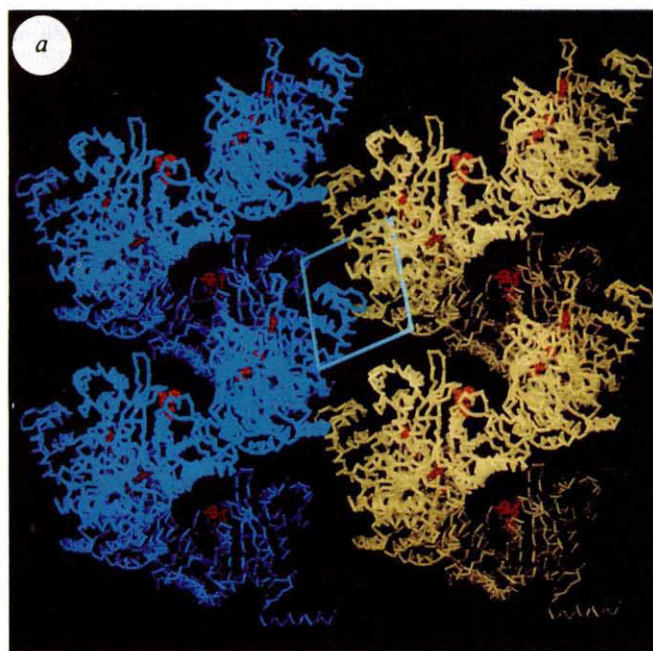


FIG. 6 Packing between filaments in the crystal. *a*, An α -carbon backbone representation of two turns each of two filaments are shown in yellow and blue, ADP in red. The boxed region shows the interfilament contact region enlarged in *b*. *b*, Close-up of the interaction between the carboxy-terminal domain of one filament (blue) and the amino end of a second filament (yellow), showing residues whose mutation affects *recA* activity. Residues 38 and 298 are described in the text. Mutations at 39 (Thr to Ile: *recA1235*), 179 (Ala to Val: *recA1212*), 184 (Gln to Lys: *recA1202*) and 301 (Gly to Asp: *recA1206*; Gly to Ser: *A1601*) lead to coprotease constitutive phenotypes; those at 301 are also recombination deficient⁵³. Mutation of Val 37 to Met (*recA803*) leads to a mutant protein with enhanced strand-exchange activity *in vitro*⁶⁷. *c*, Scheme depicting how mutations at the interpolymer interface might show a coprotease constitutive phenotype and activate *recA* protein in the absence of an inducing signal such as DNA damage. Mutations could act at steps 1, 2 or 3 to shift the equilibria shown towards the formation of active *recA* filaments. We suggest that mutations at the bundle interface (*b*) could act at step 1 to increase the effective concentration of monomers or oligomers. Such increased concentrations could permit nucleation and filament growth on 'unnatural' substrates.

(Gly 229 → Ser)⁴⁹ are found on the surface, at one edge of this 'notch'. RecA 1734 permits cleavage of λ and λ lexA repressors, but not ϕ_{80} repressor or $umuD$ (ref. 48); $recA91$ also prevents ϕ_{80} cleavage but permits λ repressor proteolysis⁴⁹. Additionally, this region is of appropriate size to accommodate a domain of a small protein, and involves residues from two independent monomers, which is consistent with the necessity of filament formation for cleavage.

Interactions between polymers

The locations of certain mutations that affect the coprotease and DNA-binding activities of $recA$ protein imply that the packing of $recA$ polymers in side-by-side parallel arrays to generate the crystal may be biologically relevant. To form this packing the carboxy terminal domain of one polymer makes contact with a region near the amino-terminus of a molecule in another polymer (Fig. 6a, b). This interface is characteristic of contacts found in crystal packing and does not resemble the tight packing and extensive complementarity found in the interface between subunits of the polymer. Surprisingly, however, in or near this interface between polymers there are a number of residues whose mutation affects the activities of $recA$ protein, even though they are far from the sites for DNA and ATP binding²⁶. Even more intriguing is the fact that all of these mutations actually improve one or more $recA$ functions, in many cases leading to phenotypes that include constitutive coprotease activity.

The most well-studied mutation in this interpolymer contact region is the $recA441$ or $Tif-1$ mutation (Glu 38 → Lys and Ile 298 → Val)⁵⁰. *In vivo*, this double mutation leads to constitutive induction of the coprotease function (and thus the SOS response), especially at 42 °C (ref. 51). *In vitro*, mutant protein competes more effectively with single-stranded DNA-binding protein for DNA, apparently as a result of faster kinetics of DNA binding⁵². The two components of this mutation are distant in an isolated monomer, yet are brought close together in the polymer-polymer interface observed in this crystal (Fig. 6b). The single Glu 38 → Lys change has been isolated ($recA1211$)⁵³ and has similar properties, without the temperature-dependence. This implies that the Ile 298 → Val change may be considered a temperature-sensitive suppressor of the first change⁵³. Both the mutant $recA1211$ protein and $recA1202$ (Gln 184 → Lys) bind ribosomal RNA and transfer RNA as substrates *in vitro* in the coprotease reaction, in addition to single-stranded DNA⁵⁴.

The phenotypes of the $Tif-1$ and related mutant proteins can be explained if we assume: (1) that bundles of $recA$ polymers exist normally *in vivo*; (2) that the $recA$ bundles are inactive in DNA binding and coprotease activities, and (3) that mutations at the interpolymer interface disrupt the normal formation of bundles of $recA$ polymers *in vivo*. That is, the phenotypes of many mutant $recA$ proteins can be explained if we assume that contacts between $recA$ monomers in the crystal, both intra-polymer and interpolymer, are biologically relevant and not just an accident of crystallization.

Several lines of evidence are consistent with the formation of bundles of $recA$ polymers in solution and their reduced ability to bind DNA. Bundles of $recA$ polymers have been observed by electron microscopy under mild conditions such as in solutions containing Mg^{2+} or spermidine at mM concentrations, or at high protein concentration^{18,55}. In fact, bundling has been observed under conditions typical of *in vitro* strand-exchange reactions⁵⁶. It has been shown that Mg^{2+} -dependent bundle formation competes with the ability of $recA$ protein to bind DNA *in vitro*^{18,57}. Consistent with the idea that the carboxy terminus promotes formation of bundles incapable of DNA binding is the observation that a proteolytic fragment of $recA$ protein lacking ~50 residues at the carboxy terminus is fully functional: it binds duplex DNA better than intact $recA$ protein and promotes strand exchange more efficiently⁵⁸. This $recA$ protein fragment does not, however, show Mg^{2+} -dependent bundle formation⁵⁸.

The function of bundle formation could be to suppress filament formation on inappropriate substrates such as duplex DNA, RNA, and single-stranded regions at replication forks, and so prevent inappropriate initiation of the SOS pathway (Fig. 6c). The $recA$ protein is highly overexpressed upon SOS induction (1 to 2% of total cell protein) and levels are still significant (~1,200 molecules per cell)⁵⁹ even when uninduced, perhaps being sufficient for bundle formation. Filament formation and induction of the SOS response are desirable only with the proper signal, presumably the formation of single-stranded DNA. By disrupting a competing reaction, mutations may be activating the coprotease activity by permitting binding to such unnatural substrates.

An alternative hypothesis is that mutations in the interpolymer interface could promote some sort of functionally important interaction between $recA$ filaments⁶⁰; bundles of $recA$ /DNA filaments have also been observed⁶¹. But the enhanced strand-exchange properties of the $recA$ proteolytic fragment⁵⁸ lacking one half of the ordered carboxy-terminal domain argues against such an interpretation.

Conveniently, for our purposes, $recA$ protein has evolved as an approximately 6₁ helical filament which is compatible with the packing in a P6₁ crystal. We speculate that the same characteristics of filament structure that evolved to regulate the DNA binding of $recA$ protein have allowed us to grow and visualize crystals of this polymer. For proteins that function as large aggregates such as filaments or bundles of filaments, the structure of the crystalline state in which important interactions are preserved is far more informative than structural studies of the monomer in solution. □

Received 16 September; accepted 20 November 1991.

- Clark, A. J. & Margulies, A. D. *Proc. natn. Acad. Sci. U.S.A.* **53**, 451-459 (1965).
- Sancar, A., Stachelek, C., Konigsberg, W. & Rupp, W. D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2611-2615 (1980).
- Hori, T., Ogawa, T. & Ogawa, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 313-317 (1980).
- Roca, A. I. & Cox, M. M. *Crit. Rev. Biochem. Molec. Biol.* **25**, 415-456 (1990).
- Kowalczykowski, S. C. *A. Rev. Biophys. biophys. Chem.* **20**, 539-575 (1991).
- Radding, C. M. *J. biol. Chem.* **266**, 5355-5358 (1991).
- Register, J. C. & Griffith, J. D. *J. biol. Chem.* **260**, 12308-12312 (1985).
- Pugh, B. F. & Cox, M. M. *J. biol. Chem.* **262**, 1326-1336 (1987).
- McEntee, K., Weinstock, G. M. & Lehman, I. R. *J. biol. Chem.* **256**, 8835-8844 (1981).
- DasGupta, C., Wu, A. M., Kahn, R., Cunningham, R. P. & Radding, C. M. *Cell* **25**, 507-516 (1981).
- Little, J. W. & Mount, D. W. *Cell* **29**, 11-22 (1982).
- Nohmi, T., Battista, J. R., Dodson, L. A. & Walker, G. C. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1816-1820 (1988).
- Little, J. W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1375-1379 (1984).
- Egelman, E. H. & Stasiak, A. *J. molec. Biol.* **191**, 677-697 (1986).
- Yu, X. & Egelman, E. H. *Biophys. J.* **57**, 555-566 (1990).
- Stasiak, A. & DiCapua, E. *Nature* **299**, 185-186 (1982).
- DiCapua, E., Engel, A., Stasiak, A. & Koller, T. *J. molec. Biol.* **157**, 87-103 (1982).
- Brenner, S. L., Zlotnick, A. & Griffith, J. D. *J. molec. Biol.* **204**, 959-972 (1988).
- DiCapua, E., Schnarr, M., Ruigrok, R. W. H., Lindner, P. & Timmins, P. A. *J. molec. Biol.* **214**, 557-570 (1990).
- Williams, R. C. & Spengler, S. J. *J. molec. Biol.* **187**, 109-118 (1986).
- Pugh, B. F. & Cox, M. M. *J. biol. Chem.* **263**, 76-83 (1988).
- DiCapua, E., Ruigrok, R. W. H., Timmins, P. A. *J. struct. Biol.* **104**, 91-96 (1990).
- McKay, D. B., Steitz, T. A., Weber, I. T., West, S. C. & Howard-Flanders, P. *J. biol. Chem.* **255**, 6662 (1980).
- Rould, M. A., Perona, J. J., Soll, D. & Steitz, T. A. *Science* **246**, 1135-1142 (1989).
- Hol, W. G. J., van Duijn, P. T. & Berendsen, H. J. C. *Nature* **273**, 443-446 (1978).
- Story, R. M. & Steitz, T. A. *Nature* **355**, 374-376 (1992).
- Blanc, M. A. *et al.* *Cold Spring Harbor Symp. quant. Biol.* **49**, 507-511 (1983).
- Rossman, M. G., Moras, D. & Olsen, K. W. *Nature* **250**, 194-199 (1974).
- Kobayashi, N., Knight, K. & McEntee, K. *Biochemistry* **26**, 6801-6810 (1987).
- Lee, B. & Richards, F. M. *J. molec. Biol.* **55**, 379-400 (1971).
- Miller, S., Lesk, A. M., Janin, J. & Chothia, C. *Nature* **328**, 834-836 (1987).
- Stasiak, A., Egelman, E. H. & Howard-Flanders, P. *J. molec. Biol.* **202**, 659-662 (1988).
- Yarranton, G. T. & Sedgwick, S. G. *Molec. gen. Genet.* **185**, 99-104 (1982).
- Müller, B., Koller, T. & Stasiak, A. *J. molec. Biol.* **212**, 97-112 (1990).
- Zlotnick, A., Mitchell, R. S. & Brenner, S. L. *J. biol. Chem.* **265**, 17050-17054 (1990).
- Kawashima, H., Hori, T., Ogawa, T. & Ogawa, H. *Molec. gen. Genet.* **193**, 288-292 (1984).
- Morand, P. F., Bianco, M. & Devoret, R. *J. Bact.* **131**, 572-582 (1977).
- Menetski, J. P. & Kowalczykowski, S. C. *J. molec. Biol.* **211**, 845-855 (1990).
- Cazaux, C., Larmarin, F. & Defais, M. *Biochimie* **73**, 281-284 (1991).
- Wierenga, R. K., De Maeyer, M. C. H. & Hol, W. G. J. *Biochemistry* **24**, 1346-1357 (1985).
- West, S. C., Cassuto, E., Mursallim, J. & Howard-Flanders, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2569-2573 (1980).
- Bryant, F. R. & Lehman, I. R. *J. biol. Chem.* **261**, 12988-12993 (1986).
- Bryant, F. R. *J. biol. Chem.* **263**, 8716-8723 (1988).
- Muench, K. A. & Bryant, F. R. *J. biol. Chem.* **266**, 844-850 (1991).
- Wang, W.-B. & Tessman, E. S. *J. Bact.* **168**, 901-910 (1986).
- Zlotnick, A. & Brenner, S. L. *J. molec. Biol.* **209**, 447-457 (1989).
- de Jong, E. A. M. *et al.* *J. molec. Biol.* **206**, 133-152 (1989).
- Duttreix, M. *et al.* *J. Bact.* **171**, 2415-2423 (1989).

49. Ogawa, H. & Ogawa, T. *Adv. Biophys.* **21**, 135–148 (1986).
 50. Knight, K. L., Aoki, K. H., Ujita, E. L. & McEntee, K. *J. biol. Chem.* **259**, 11279–11283 (1984).
 51. Kirby, E. P., Jacob, F. & Goldthwait, D. A. *Proc. natn. Acad. Sci. U.S.A.* **58**, 1903–1910 (1967).
 52. Lavery, P. E. & Kowalczykowski, S. C. *J. molec. biol.* **203**, 861–874 (1988).
 53. Wang, W.-B. & Tessman, E. S. *J. Bact.* **168**, 901–910 (1986).
 54. Wang, W.-B., Tessman, E. S. & Tessman, I. *J. Bact.* **170**, 4823–4827 (1988).
 55. Griffith, J. & Shores, C. G. *Biochemistry* **24**, 158–162 (1985).
 56. Wilson, D. H. & Benight, A. S. *J. biol. Chem.* **265**, 7351–7359 (1990).
 57. Morricat, S. W. & Cox, M. M. *Biochemistry* **24**, 760–767 (1985).
 58. Benedict, R. C. & Kowalczykowski, S. C. *J. biol. Chem.* **263**, 15513–15520 (1988).
 59. Salles, B. & Paoletti, C. *Proc. natn. Acad. Sci. U.S.A.* **80**, 65–69 (1983).
 60. Lauder, S. D. & Kowalczykowski, S. C. *J. biol. Chem.* **266**, 5450–5458 (1991).

61. Egelman, E. H. & Stasiak, A. *J. molec. biol.* **200**, 329–349 (1988).
 62. Sancar, A. & Rupp, W. D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3144–3148 (1979).
 63. Wang, B. C. *Meth. Enzym.* **115**, 90–112 (1985).
 64. Brunger, A. T. *Program X-PLOR Version 2.1* (Yale Univ., Connecticut, 1990).
 65. Priestle, J. P. *J. appl. Crystallogr.* **21**, 572–576 (1988).
 66. Weismann, J. M. & Weinstock, G. M. *DNA* **7**, 389–398 (1988).
 67. Madiraju, M. V. S., Templin, A. & Clark, A. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6592–6596 (1988).

ACKNOWLEDGEMENTS. This work is dedicated to the memory of Paul Howard-Flanders. We thank members of the structural biology group at Yale, particularly J. Friedman and S. Schultz. This research was supported by a NIH grant to T.A.S. and a NSF predoctoral fellowship to R.M.S. The full coordinates have been sent to the Protein Data Base at Brookhaven.

LETTERS TO NATURE

An observational test for the existence of a planetary system orbiting PSR1257+12

F. A. Rasio, P. D. Nicholson, S. L. Shapiro & S. A. Teukolsky

Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA

FOLLOWING the first report¹ of an object of planetary mass orbiting a pulsar, Wolszczan and Frail² have now reported the even more surprising discovery of two planet-size companions in orbit around the nearby millisecond pulsar PSR1257+12. The orbital periods of the two planets are about 98 days and 67 days, very close to a 3:2 ratio. Here we point out that, because of this near commensurability, the mutual gravitational perturbations of the two planets should produce not only small secular changes, but also larger periodic changes in their orbital elements. In particular, we find that changes in the eccentricities and orbital periods should become measurable within a few years. Such a measurement would help determine the three masses in the system and the inclinations of the orbits. More importantly, a detection of these changes, if they accord with the theoretical predictions presented here, would provide irrefutable confirmation that the periodic residuals observed by Wolszczan and Frail are indeed caused by orbiting planets, rather than some other effect. For the single planet-size object previously reported¹ around the pulsar PSR1829–10, there is no dynamical test analogous to the one proposed here to confirm the planetary interpretation.

To each planet orbiting the pulsar there corresponds a mass function³ $f_j = (m_j \sin i)^3 / (M + m_j)^2$. Here M is the mass of the central neutron star, m_j is the mass of the planet, with $j = 1$ for the inner planet and $j = 2$ for the outer planet, and i is the angle between the orbital plane and the plane of the sky (we assume that the orbital plane is the same for both planets; see below). From the observed parameters of the system, we find $f_1 = 5.45 \times 10^{-16} M_\odot$ and $f_2 = 3.12 \times 10^{-16} M_\odot$. We can safely neglect terms of order $m_j/M \sim 10^{-5}$ in the above expression for the mass function. We deduce that $m_1 = 3.4 M_\oplus M_{1.4}^{2/3} (\sin i)^{-1}$ and $m_2 = 2.8 M_\oplus M_{1.4}^{2/3} (\sin i)^{-1}$, where $M_{1.4}$ is the neutron star mass in units of the canonical value of $1.4 M_\odot$, and $M_\oplus$ is the mass of the Earth. The corresponding semimajor axes in AU are $a_1 = 0.36 M_{1.4}^{1/3}$ and $a_2 = 0.47 M_{1.4}^{1/3}$. The other orbital elements are directly measurable, independent of M and $\sin i$. The values currently observed² are $e_1 = 0.022$ and $e_2 = 0.020$ for the eccentricities, and $\omega_1 = 252^\circ$ and $\omega_2 = 107^\circ$ for the longitudes of pericentre (measured from the ascending node).

Using these orbital elements, and the times of pericentre passage given by Wolszczan and Frail², one can construct a complete set of initial data for the system⁴. The dynamical evolution of the system is then determined by solving Newton's equations for three point masses interacting gravitationally. We have done a high-accuracy numerical integration of these equations, using the Bulirsch–Stoer algorithm⁵. We make the

problem dimensionless by setting $G = a_1 = m_1 = 1$. The numerical solution then depends on only one parameter, the ratio M/m_1 . The results corresponding to $M/m_1 = 1.4 \times 10^5$ are illustrated in Fig. 1. These results are little changed if other, more distant planets are present in the system. In particular, we have verified that the presence of a third planet with $m_3 \approx 1 M_\oplus$ and $a_3 \approx 1$ AU (for which there is preliminary evidence²) has practically no influence on the results shown in Fig. 1. We find that, on a timescale of ~ 10 yr, any orbital element w closely follows an evolution of the type

$$w(t) = A + Bt + C \sin(Dt + E) \quad (1)$$

The linear term is only apparent in the evolution of ω_j , for which $B \approx 0.04^\circ \text{ yr}^{-1}$. The evolution of all orbital elements is dominated by the periodic term. The largest fractional change is that of the eccentricity of the outer planet, for which $C \approx 4 \times 10^{-4}$, corresponding to $\Delta e_2/e_2 \approx 2 \times 10^{-2}$. The period is $2\pi/D \approx 5.5$ yr for all elements. The constants A and E are determined by the initial conditions.

We can gain a better physical understanding of these results by examining Lagrange's equations for the system (see ref. 6 for a general discussion and refs 7 and 8 for a clear summary and useful Solar System examples). The disturbing function for each planet, due to its gravitational perturbation by the other, can be separated into two parts. The secular part is made of

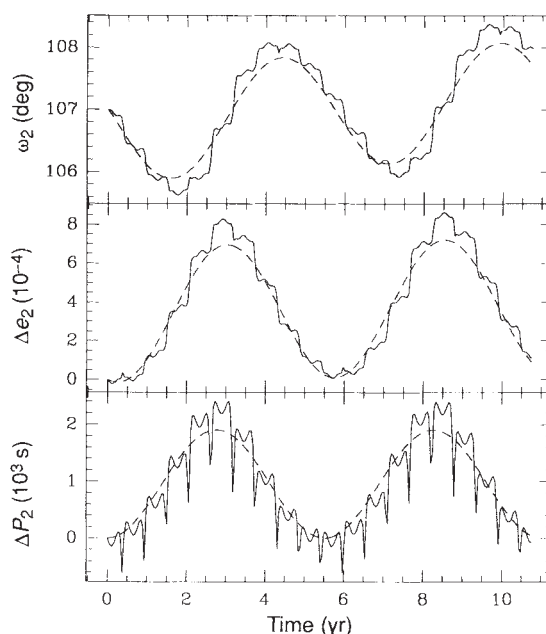


FIG. 1 Orbital evolution of the PSR1257+12 system. The solid lines are from a direct numerical integration of Newton's equations for three point masses. The dashed lines are from perturbation theory (equations (2) and (6)). The instantaneous (osculating) values of the longitude of pericentre, and of the changes in eccentricity and orbital period of the outer planet, are shown as a function of time. The orbital elements of the inner planet follow a similar evolution, but with somewhat smaller amplitudes.

TABLE 1 Secular evolution parameters

	$i=1$ eigenmode	$i=2$ eigenmode
g_i (deg yr $^{-1}$)	4.39×10^{-2}	2.93×10^{-3}
$2\pi/g_i$ (yr)	8,210	123,000
E_{1i}	-1.94×10^{-2}	-6.52×10^{-3}
E_{2i}	$+2.06 \times 10^{-2}$	-6.50×10^{-3}
β_i (deg)	88.7	13.2

terms that do not depend on the mean longitudes of either planet. In addition, the existence of a near 3:2 commensurability in the mean motions of the two planets implies that those terms in the disturbing function that depend on the particular combination of mean longitudes $2\lambda_1 - 3\lambda_2$ will contribute large, periodic changes. We treat each contribution in turn and obtain simple analytical expressions for the coefficients of equation (1).

Following Brouwer and Clemence⁶, we write the solution of Lagrange's equations corresponding to the secular part of the disturbing function as

$$\begin{aligned} h_j &\equiv e_j \sin \omega_j = \sum_{i=1,2} E_{ji} \sin(g_i t + \beta_i) \\ k_j &\equiv e_j \cos \omega_j = \sum_{i=1,2} E_{ji} \cos(g_i t + \beta_i) \end{aligned} \quad (2)$$

There is no secular variation of the semimajor axes or orbital periods. The frequencies g_i and the amplitudes E_{ji} can be calculated as the eigenvalues and eigenvector components of a 2×2 matrix with coefficients given explicitly in terms of the orbital elements⁷. The norms of the eigenvectors and the phases β_i can be calculated from a set of initial conditions. The results for the PSR1257+12 system, assuming $M = 1.4 M_\odot$ and $\sin i = 1$, are given in Table 1. We have numbered the two eigenstates so that $g_1 > g_2$. We find that the current state of the system is close to a pure g_1 eigenstate with $E_{11} \approx -E_{21}$. This corresponds to $e_j \approx |E_{ji}| = \text{constant}$, $\omega_1 \approx g_1 t + \beta_1 + \pi$, and $\omega_2 \approx \omega_1 - \pi$. A similar eigenmode structure is observed for the two outer satellites of Uranus, Titania and Oberon, which are also in a near 3:2 resonance^{7,8}. By comparing equations (2) and (1) for ω_j , we see that $B \approx g_1$. This can be calculated explicitly as

$$g_1 = \frac{1}{8} n_1 \frac{m_1}{M} \alpha b_{3/2}^{(1)}(\alpha) \{q\alpha + \nu + [(q\alpha - \nu)^2 + 4q\alpha\nu\beta^2]^{1/2}\} \quad (3)$$

In this expression, n_i is the mean orbital motion, the $b_{3/2}^{(n)}$ are Laplace coefficients⁶, and we have defined the non-dimensional ratios $\alpha = a_1/a_2$, $q = m_2/m_1$, $\nu = n_2/n_1$, and $\beta = b_{3/2}^{(2)}(\alpha)/b_{3/2}^{(1)}(\alpha)$.

We now turn to the near-resonant terms in the disturbing function. For the 3:2 case, there are two such terms, which depend on the mean longitudes as $\cos \phi_j$ with $\phi_j = 2\lambda_1 - 3\lambda_2 + \omega_j$. The corresponding amplitudes are given explicitly in terms of the orbital elements⁶. To lowest order in e , we find that Lagrange's equations reduce to

$$\dot{\omega}_j = F_j(\alpha) \frac{n_j}{e_j} \frac{m_k}{M} \cos \phi_j, \quad \dot{e}_j = F_j(\alpha) n_j \frac{m_k}{M} \sin \phi_j \quad (4)$$

and a similar expression for $\dot{a}_j$, with $k = 2(1)$ if $j = 1(2)$, and

$$\begin{aligned} F_1(\alpha) &= -\frac{1}{2} \left(6\alpha + \alpha^2 \frac{d}{d\alpha} \right) b_{1/2}^{(3)}(\alpha) \\ F_2(\alpha) &= \frac{1}{2} \left(5 + \alpha \frac{d}{d\alpha} \right) b_{1/2}^{(2)}(\alpha) \end{aligned} \quad (5)$$

To solve equations (4) analytically, we use the fact that $|\dot{\omega}_j| \ll |2n_1 - 3n_2|$ and we keep the orbital elements constant at their present values when evaluating the right-hand sides. (These approximations are justified for $t \ll 2\pi/|\dot{\omega}_j|$ and as long as the perturbations are small.) We can then write $\phi_j =$

$(2n_1 - 3n_2)t + \phi_j^{(0)}$, where $\phi_j^{(0)}$ is the initial value of ϕ_j , and find

$$\omega_j(t) = \frac{F_j(\alpha)}{e_j} \frac{m_k}{M} \frac{n_j}{(2n_1 - 3n_2)} \{ \sin [(2n_1 - 3n_2)t + \phi_j^{(0)}] - \sin \phi_j^{(0)} \} \quad (6)$$

and similar expressions for $e_j(t)$ and the osculating orbital period $P_j(t)$. These analytical solutions are shown in Fig. 1 for the outer planet (dashed lines). By comparing them with equation (1), we obtain simple expressions for the coefficients C and D . In particular we have $D = 2n_1 - 3n_2$ for all orbital elements.

The rapid fluctuations apparent in the numerical solution (Fig. 1) are caused by all the nonresonant and nonsecular terms in the disturbing function, which have been neglected in our analytical treatment. The largest such fluctuations are those corresponding to close encounters between the two planets, which occur roughly every $3 \times (2\pi/n_1) \approx 200$ days. In particular, these close encounters produce changes in the instantaneous (osculating) orbital periods of amplitude $\Delta P/P \approx 10^{-4}$, occurring over a timescale of ~ 1 month. The current error bars on the orbital periods² (which are measured by fitting data over a period of time $\gg 1$ month) are $\sim 10^{-4}$, indicating that these effects may be marginally detectable already.

The amplitudes of both secular and periodic changes in the orbital elements are all proportional to the ratio m_i/M , which scales as $M^{-1/3}(\sin i)^{-1}$. Therefore, a measurement of any of these changes will provide a constraint on the combination $M^{1/3} \sin i$, but will not allow M and $\sin i$ to be determined independently. If the inclinations of the two orbits are slightly different, a measurement of any orbital change for both planets will provide independent estimates of $M^{1/3} \sin i_1$ and $M^{1/3} \sin i_2$. Our analysis remains valid as long as the relative inclination $\delta i = i_1 - i_2$ is small. This is because near-resonant periodic terms involving the inclinations and node longitudes do not arise to lowest order in δi for the 3:2 case⁸. Secular variations in these quantities do arise, but, in the limit of small eccentricities and small relative inclination, they are completely decoupled from those involving eccentricities and longitudes of pericentre⁷. Should δi turn out to be large, a more general analysis would be necessary.

A strict upper limit on the masses of the two planets can also be obtained by considering the long-term dynamical stability of the system. Stability requires that the minimum separation between the two planets always remain larger than the radius of their 'Hill's spheres'. This condition can be written approximately⁹ as $\Delta a/a_j > 2.1(m_j/M)^{1/3}$. As $m_1 \approx m_2$, this implies for both planets $M/m_j > 7 \times 10^2$, or $m_j/M_\oplus < 6 \times 10^2 M_{1.4}$. By numerically integrating the equations of motion of the system over a timescale of $\sim 10^4$ yr for many different values of the ratio M/m_1 , we find a stability condition $M/m_1 > 8 \times 10^2$, in good agreement with the simple analytical estimate.

Quite apart from the determination of the masses and orbital inclinations, our results provide a critical test for the basic interpretation of the radio observations. The measured orbital elements of the two planets must all be evolving exactly as described above. A confirmation of this evolution would provide irrefutable evidence that a planetary system outside our own Solar System has indeed been found. $\square$

- Bailes, M., Lyne, A. G. & Shemar, S. L. *Nature* **352**, 311–313 (1991).
- Wolszczan, A. & Frail, D. A. *Nature* **355**, 145–147 (1992).
- Shapiro, S. L. & Teukolsky, S. A. *Black Holes, White Dwarfs, and Neutron Stars* (Wiley, New York, 1983).
- Danby, J. M. A. *Fundamentals of Celestial Mechanics* (Willmann-Bell, Richmond, 1988).
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes* (Cambridge University Press, Cambridge, 1986).
- Brouwer, D. & Clemence, G. M. *Methods of Celestial Mechanics* (Academic, New York, 1961).
- Dermott, S. F. & Nicholson, P. D. *Nature* **319**, 115–120 (1986).
- Malhotra, R., Fox, K., Murray, C. D. & Nicholson, P. D. *Astr. Astrophys.* **221**, 348–358 (1989).
- Gladman, B. & Duncan, M. *Astr. J.* **100**, 1680–1691 (1990).

ACKNOWLEDGEMENTS. We thank A. Wolszczan for discussions and for communicating results before publication. We also thank J. Lissauer for drawing our attention to the Hill's sphere constraint. This work has been supported in part by NSF and NASA grants to Cornell University.

High-quality superconducting thin films of $\text{Bi}_2\text{Sr}_2\text{Ca}_2\text{Cu}_3\text{O}_x$ grown *in situ* by metalorganic CVD

K. Endo, H. Yamasaki, S. Misawa, S. Yoshida & K. Kajimura

Electrotechnical Laboratory, 1-1-4 Umezono, Tsukuba, Ibaraki 305, Japan

THE superconducting copper oxide system Bi-Sr-Ca-Cu-O (BSCCO) has attracted much attention owing to its high transition temperatures T_c and its stability on exposure to air, relative to the corresponding properties of the Y-Ba-Cu-O system. But despite intense effort, growth techniques capable of yielding as-deposited films with high T_c and high critical current density, J_c , have not yet been forthcoming. Here we report the preparation and properties of *c*-axis-oriented single-phase $\text{Bi}_2\text{Sr}_2\text{Ca}_2\text{Cu}_3\text{O}_x$ films grown *in situ* without post-annealing by metalorganic chemical vapour deposition (MOCVD). Our films have the highest T_c and J_c at 77 K reported so far for BSCCO films grown *in situ*; one film had a zero-resistance T_c of 97 K, and a J_c of $3.8 \times 10^5 \text{ A cm}^{-2}$ at 77 K in zero magnetic field. This film showed little degradation of J_c in magnetic fields of up to 8 T applied parallel to the crystallographic *a*-*b* plane. The magnetic-field dependence of J_c and observed surface morphologies strongly suggest that these as-grown films have no weak links, which is a promising characteristic for device applications.

The BSCCO system has many superconducting phases with various values of T_c : examples of bulk materials are the high- T_c phase $\text{Bi}_2\text{Sr}_2\text{Ca}_2\text{Cu}_3\text{O}_x$ ('2223') with $T_c \approx 110 \text{ K}$ and the low- T_c phase $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_x$ ('2212') with $T_c \approx 80 \text{ K}$. It is difficult, however, to obtain films of the high- T_c phase 2223. Most of the films reported so far consist of the low- T_c 2212 phase in an as-deposited state¹⁻³. No growth technique has yet produced films of 2223 with high T_c (>90 K) without post-annealing.

The MOCVD method we have adopted is much closer to an equilibrium process than are other film-preparation methods such as molecular beam epitaxy and sputtering. In previous work⁴, we succeeded in growing single-phase $\text{Bi}_2\text{Sr}_2\text{Ca}_2\text{Cu}_3\text{O}_x$ films on MgO substrates *in situ* without post-annealing by MOCVD. For these films, however, T_c was only 74 K. *In situ* lead doping during growth using bis(-2,2,6,6-tetramethyl-3,5-heptanedione)-lead, $\text{Pb}(\text{DPM})_2$, did not increase T_c much, although a Pb-doped high- T_c phase was formed⁵. Therefore, we speculated that the growth condition was still far from equilibrium, and this was confirmed when we found that T_c increased with increasing film thickness and with decreasing growth rate⁶.

Here we used a cold-wall type MOCVD apparatus with a horizontal reaction tube. The source materials were triphenyl bismuth $[\text{Bi}(\text{C}_6\text{H}_5)_3]$, $\text{Sr}(\text{DPM})_2$, $\text{Ca}(\text{DPM})_2$ and $\text{Cu}(\text{DPM})_2$. They were charged in individual vaporizers, heated to appropriate temperatures depending on the materials and carried into a quartz reactor using argon gas with a flow rate of $770 \text{ cm}^3 \text{ min}^{-1}$. Gas lines were heated to temperatures higher than those of the

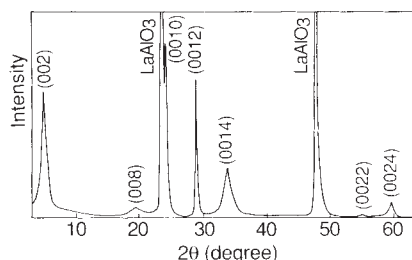


FIG. 1 X-ray diffraction pattern of a BSCCO film grown *in situ* by MOCVD.

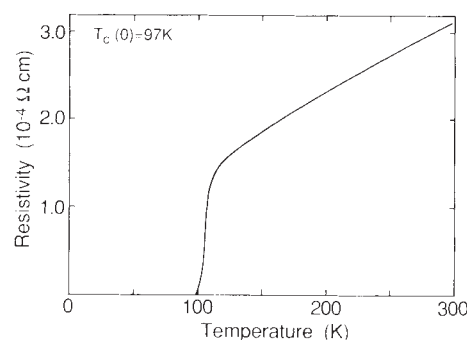


FIG. 2 Temperature dependence of resistivity for the as-grown BSCCO film shown in Fig. 1. $T_c(0)$ is 97 K.

vaporizers to prevent condensation of source materials. Oxygen gas was separately fed into the reactor at a flow rate of $675 \text{ cm}^3 \text{ min}^{-1}$. Substrates of $\text{LaAlO}_3(100)$ single crystals, $5 \times 10 \times 0.5 \text{ mm}^3$ in size and having very small lattice mismatches with BSCCO, were placed on an Inconel susceptor and then inductively heated to 800°C , above which the bismuth content in the film abruptly decreased and the film surface was roughened appreciably. The total pressure in the reactor was kept at 50 torr during the film growth, and the oxygen partial pressure was 23 torr. After growth, the supply of argon gas to the source materials was cut off and then the films were cooled down to less than 300°C in 10 min by switching off the heater without changing the oxygen pressure. The growth rate was $\sim 12 \text{ Å h}^{-1}$. Such an extremely low growth rate is admittedly a disadvantage for industrial applications, but it is our present purpose to demonstrate that such high-quality superconducting films can be grown *in situ* without post-annealing. The thickness of the films obtained here is 420–960 Å. The grown films had mirror-like shiny surfaces. As shown in Fig. 1, all the X-ray diffraction peaks of as-grown films were definitely indexed as (00*n*) of the 2223-high- T_c phase; that is, the films consisted of single-phase $\text{Bi}_2\text{Sr}_2\text{Ca}_2\text{Cu}_3\text{O}_x$ with the *c*-axis oriented normal to the substrate. Surface morphology examined by scanning electron microscopy showed no apparent grain boundaries. Compositions were confirmed by inductively-coupled-plasma atomic emission spectroscopy. Resistivity and J_c measurements were done by a conventional four-probe d.c. method using evaporated gold-film pads as electrodes without making any bridge pattern in the film. The voltage tap distance was 3 mm. Critical current measurements with pulse currents were also carried out to confirm that the heating effect at the electrodes was negligible.

Figure 2 shows the temperature dependence of the resistivity of the same film as that of Fig. 1. The zero-resistance critical temperature $T_c(0)$ is 97 K, which is the highest value of which we are aware so far for as-grown BSCCO films. The values of $T_c(0)$ for the 2223-high- T_c single-phase films prepared here were 94–97 K.

Figure 3a shows the magnetic field dependence of J_c of the film prepared in the same batch as that shown in Fig. 1 measured at 77 K. The magnetic field H was applied parallel to the film surface within an accuracy of less than one degree and perpendicular to the direction of measuring currents. In zero magnetic field J_c is $3.8 \times 10^5 \text{ A cm}^{-2}$, with a criterion of $2 \mu\text{V cm}^{-1}$ electric field used to determine J_c . The criterion is important if J_c values are to be compared. This value of J_c is the highest reported so far for BSCCO films. J_c decreased very slightly in low magnetic field (0–0.1 T), and no hysteresis was observed in the J_c - H curve. These facts imply that there are no signs of weak links in these films, and this is consistent with the electron microscope observation. The J_c - H curve is so flat that the values of J_c in fields of 1 and 8 T are 3.3×10^5 and $9.1 \times 10^4 \text{ A cm}^{-2}$, respectively, when the magnetic field is perpendicular to the measuring current. These J_c values are comparable to those of YBCO films

in a magnetic field >1 T, although still an order of magnitude lower than YBCO in zero magnetic field^{7,8}. Further evidence for the weak-link-free J_c behaviour is obtained from measurements taken with the magnetic field applied parallel to the current, $J_c(\parallel)$, which corresponds to the Lorentz-force-free configuration. In Fig. 3b, we compare $J_c(\parallel)$ with the usual J_c value under magnetic field perpendicular to the current, $J_c(\perp)$. The data were taken from a lower-quality sample than that of Fig. 3a, as the latter was broken after the experiment. Figure 3b shows that J_c was enhanced with $J_c(\parallel)/J_c(\perp) = 1-2$ observed in magnetic fields up to 14 T. This feature strongly suggests that J_c was determined by the balance between the flux pinning strength and the Lorentz force, and that the system does not consist of weak-link networks. In contrast, Kumakura *et al.* have reported that $J_c(\parallel)$ is essentially equal to $J_c(\perp)$ at 4.2 K under high magnetic field for textured $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_y$ tapes⁹. They interpreted the result as follows: the microscopic superconducting current takes a zig-zag path because of the random orientation of grains in the plane perpendicular to the c -axis, and the volume pinning force is independent of the field direction⁹.

In BSCCO, which is highly anisotropic¹⁰, magnetic flux lines parallel to the a - b plane can penetrate easily into insulating

block layers embedded among superconducting CuO_2 layers without appreciable increase in the free energy of superconductivity, but the flux lines can hardly move across the CuO_2 layers. Tachiki and Takahashi¹¹ therefore proposed that the insulating layers become strong pinning centres, preventing movement of flux lines (intrinsic pinning). But it has often been suggested that the intrinsic pinning in BSCCO films may not be very strong and that the insulating block layers between superconducting CuO_2 layers may not act as pinning centres, because earlier measurements of J_c in BSCCO at high temperature (~ 77 K) in a magnetic field (>1 T) have so far been much lower than that of YBCO films^{7,8}. Weak-link problems, originating from grain boundaries or other defects, may have obscured the intrinsic properties of BSCCO. Our results evidently show that weak-link-free BSCCO films have very high J_c with strong flux pinning. The effect on J_c of the angle between the magnetic field and the a - b plane of these films has been also investigated, to obtain more insight into the flux-pinning mechanism. As shown in Fig. 3c, J_c decreases abruptly from $\sim 10^5$ A cm⁻² to zero as the applied field slightly deviates from the a - b plane. This result supports the idea of intrinsic pinning, in which the pinning strength is very strong for magnetic fluxes parallel to the a - b plane, and very weak for those parallel to the c -axis. When BSCCO films with even higher T_c , comparable to the bulk value, 110 K, are properly prepared, we expect that J_c will increase considerably. □

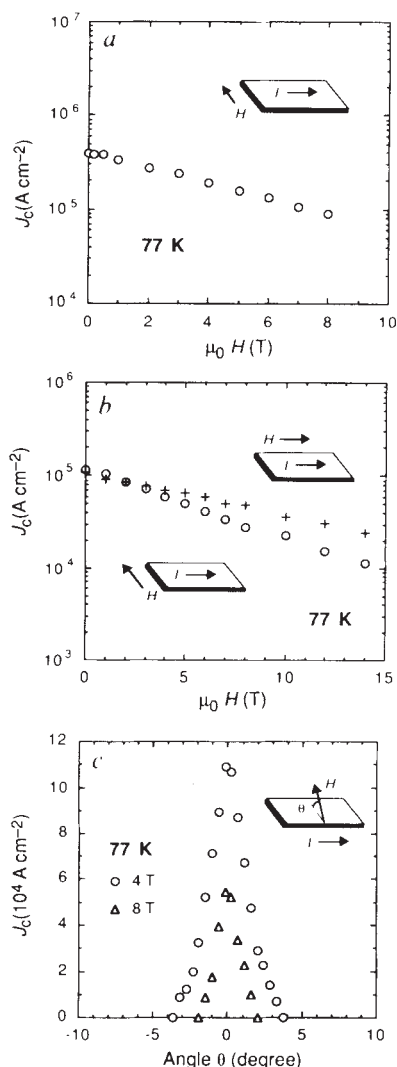


FIG. 3 a, The critical current density J_c at 77 K as a function of magnetic field applied parallel to the film surface and perpendicular to the current direction. The sample was prepared in the same batch as that in Figs 1 and 2, and had $T_c(0)$ of 95 K. b, J_c observed in another sample for magnetic fields applied both parallel and perpendicular to the current direction. c, J_c as a function of the angle between applied magnetic field and the film surface.

Received 28 November; accepted 5 December 1991.

1. Eckstein, J. N. *et al.* *J. Vac. Sci. Technol.* **B70**, 319–323 (1989).
2. Schmitt, P., Schultz, L. & Saemann-Ischenko, G. *Physica C* **168**, 475–478 (1990).
3. Matsushima, T., Ichikawa, Y., Adachi, H., Setsune, K. & Wasa, K. *Solid State Commun.* **76**, 1201–1204 (1990).
4. Endo, K. *et al.* *Jap. J. appl. Phys.* **29**, 294–297 (1990).
5. Endo, K. *et al.* in *Advances in Superconductivity II* (eds Ishiguro, T. & Kajimura, K.) 912–915 (Springer, Tokyo, 1990).
6. Endo, K., Nakatsuka, K., Misawa, S. & Yoshida, S. in *Advances in Superconductivity III* (eds Kajimura, K. & Hayakawa, H.) 1017–1020 (Springer, Tokyo, 1991).
7. Roas, B., Schultz, L. & Endres, G. *Appl. Phys. Lett.* **53**, 1557–1559 (1988).
8. Watanabe, K. *et al.* *Appl. Phys. Lett.* **54**, 575–577 (1989).
9. Kumakura, H., Togano, K., Maeda, H., Kase, J. & Morimoto, T. *Appl. Phys. Lett.* **58**, 2830–2832 (1991).
10. Iye, Y., Tamegai, T. & Nakamura, S. *Physica C* **174**, 227–232 (1991).
11. Tachiki, M. & Takahashi, S. *Solid State Commun.* **70**, 291–295 (1989).

Atomic replacement and vacancy formation and annihilation on iridium surfaces

Tien T. Tsong*† & Chong-lin Chen‡

* Institute of Physics, Academia Sinica, Nankang, Taipei 11529 China

‡ Physics Department, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

THE formation of nanometre-scale surface structures for quantum electronic devices, and the possibility of using the scanning tunnelling microscope and related instruments for atomic-scale surface modification^{1–6}, make necessary a detailed understanding of surface atomic processes. It has been observed previously^{7–11} that surface diffusion of an atom can take place by a series of exchanges with atoms in the surface layer. Here we use the field ion microscope to observe surface atomic processes when the chemical nature of the adatom differs from that of the surface atoms. On the iridium {001} surface, an adsorbed rhenium atom will displace an iridium atom from the substrate to form a Re–Ir–vacancy bound complex at about 230 K. On heating the sample to about 300 K, we find that this complex dissociates and the rhenium atom becomes

† To whom correspondence should be addressed. On leave from the Pennsylvania State University.

incorporated into the surface—a kind of atomic-scale surface alloying. Alternatively, the Re-Ir cluster can be removed by field evaporation to leave a lattice vacancy, into which another adatom can subsequently diffuse.

To study surface atomic processes^{12,13} in the field ion microscope (FIM)¹⁴, a very high vacuum, below the low 10^{-11} Torr range, is required. In our experiment, the FIM tip is first bombarded by neon ions and then degassed in ultra-high vacuum at $\sim 1,000$ K. The adatoms are deposited on the low-temperature field-evaporated surface by heating a rhenium coil. This coil is thoroughly degassed in ultra-high vacuum at the deposition temperature before use. Experimental details have been published elsewhere¹³⁻¹⁵. For the image gas, we use fused-quartz diffused helium to avoid contamination of the surface. During imaging, the tip is cooled to below 30 K. At this temperature, the image field will not induce changes in the positions of adatoms or atomic clusters. To induce a surface atomic process, the tip is pulse-heated to the desired temperature without the image field, using an electronically controlled pulsed power supply. The rise time of the temperature pulse is less than 0.2 s, and there is no overshoot.

When a Re atom is deposited on a low-temperature field-evaporated Ir {001} surface, which has the (1×1) structure, its image is unchanged, except for the additional circular image of the adatom. This adatom can be field-evaporated at a field of $\sim 4.5 \pm 0.5 \text{ V } \text{\AA}^{-1}$ without producing any observable change in the substrate. To induce an atomic process, the surface is heated in the absence of the image field for a period whose length is such that a change is observed after several such periods. It is then quenched to ~ 30 K before the image field is applied again. When the heating temperature is below 200 K, no changes are observed. Around 220 K, we occasionally observe a sudden

change from a circular to an oblong image spot, presumably produced because the Re adatom has formed a Re-Ir dimer by activating a substrate atom onto the surface, as shown in Fig. 1c. When the surface is heated further, to 210–260 K, occasional changes of the orientation of the oblong image spot can be seen (Fig. 1c and d). That a Re-Ir-vacancy complex is indeed formed can be confirmed if a lattice vacancy is created in the surface layer. On gradual low-temperature field evaporation of the surface, which first removes the cluster, followed by atoms from the plane edges, a lattice vacancy is found to be left in the top surface layer below the Re-Ir dimer. This vacancy can be seen before the receding edge of the layer reaches it, as shown in Fig. 2. Figure 2g and i illustrates the atomic processes involved.

The dimer-vacancy complex can be thermally dissociated if the temperature exceeds ~ 280 K. On dissociation, the Re atom fills the vacancy, and the displaced Ir atom disappears from the top surface layer. The latter occurs because the diffusivity of Ir adatoms on the Ir {001} surface at such a temperature is very high, and the plane boundary is non-reflective to Ir adatoms¹⁵. The substitutional Re atom shows a slightly brighter image intensity when we start removing the surface layer by field evaporation. It can be seen well before the plane edge reaches the Re atom, as shown in Fig. 3. This is consistent with the atomic-replacement adatom diffusion mechanism proposed and studied in ref. 16 and in earlier FIM experiments⁷⁻¹¹. In contrast to earlier experiments, we find that for Re on Ir {001} the atomic replacement occurs in two steps: the first is the change from Re-adatom state to the metastable dimer-vacancy state, the second is the transition to the final replaced state.

The heights of the activation barriers for these steps can be derived by measuring the temperature dependence of the average

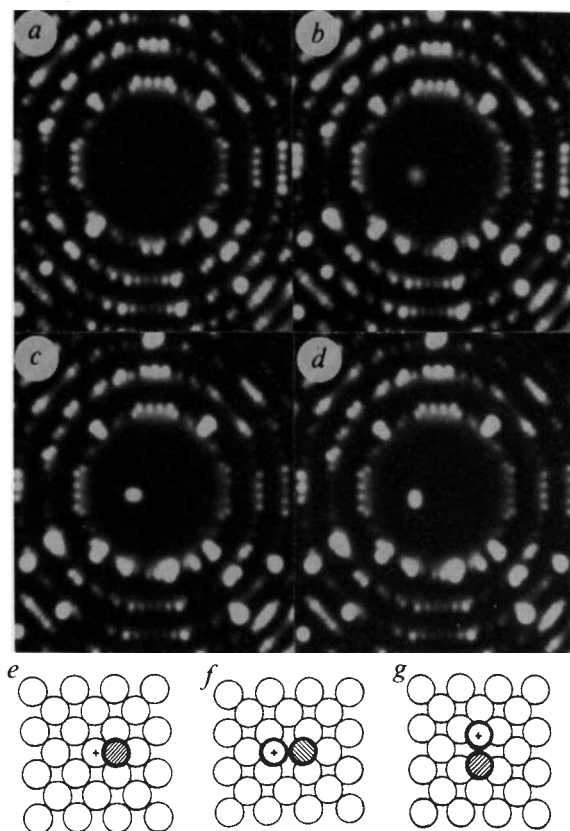


FIG. 1 a, Image of an Ir {001} surface. b, A Re atom is deposited on it, appearing as a circular spot. c, When the sample is heated to ~ 230 K, a Re-Ir dimer is formed which appears oblong-shaped. This is a typical image shape for a metal dimer on a metal surface. d, On heating to ~ 262 K, its orientation changes.

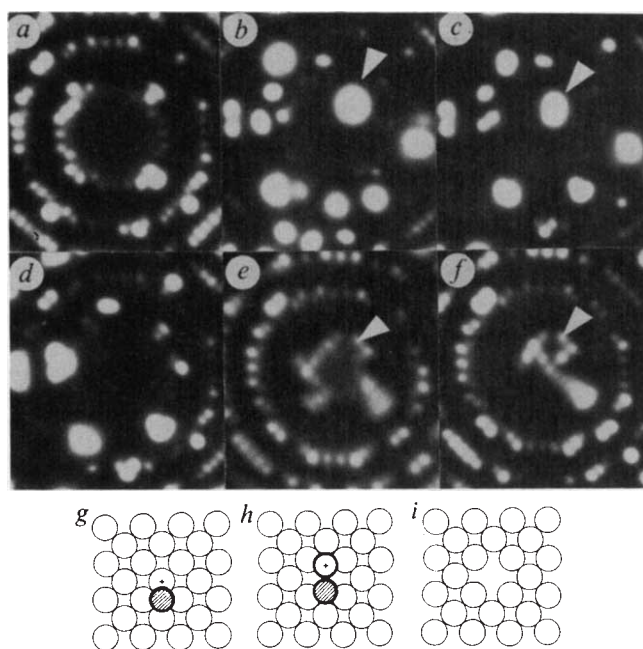


FIG. 2 a, Image of an Ir {001} surface. b, Image of a vapour-deposited Re adatom on an Ir {001} surface. For a very sharp tip, FIM images of an adatom always appear excessively bright and fat. The other bright spots are other deposited Re adatoms on terraces of the second and third layer of the Ir surface, and will not be discussed here. A Re-Ir dimer is formed by heating to 242 K, as seen in c. d, On heating to ~ 300 K, the dimer dissociated and disappeared. On gradual field evaporation of the dimer and the substrate, a lattice vacancy is found in the substrate directly beneath the dimer (e–f). Arrows indicate a Re atom in b, a Re-Ir dimer in c, and the vacancy in e and f. The positions of the dimer and vacancy are compared by digitized images using a computer. These atomic steps are illustrated in g–i.

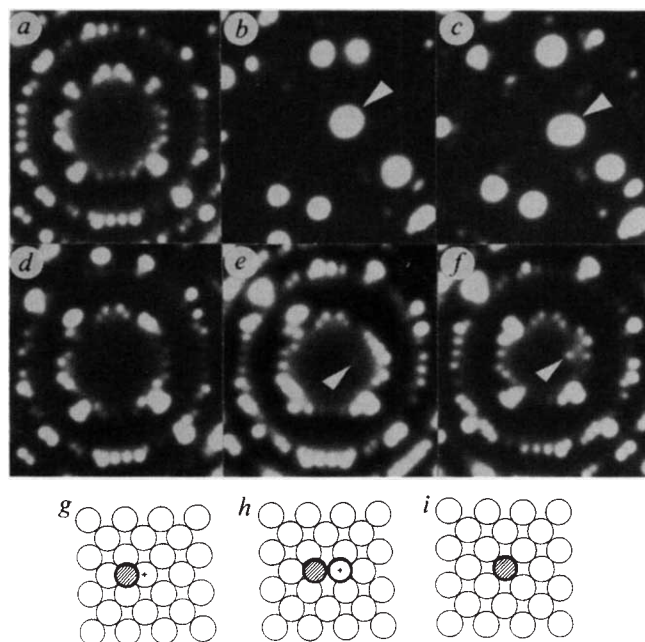


FIG. 3 After a Re-Ir-vacancy complex is formed (*a-c*; these are equivalent to *a-c* in Fig. 2), the surface is heated to 280 K again to induce thermal dissociation. Both atoms disappear from the surface (*d*) but on gradual low-temperature field evaporation, the Re atom is found to have been incorporated into the substrate lattice at the original location. The substitutional Re atom is noticeably brighter than an Ir atom in the same site would be, as seen in *e* and *f*. These atomic steps are illustrated in *g-i*.

time taken to produce the complex and that for its dissociation. The transition rate, κ , is related to the average formation time, τ , and the activation barrier height, E_b , by

$$\kappa = 1/\tau = \nu \exp(-E_b/kT)$$

where ν is the frequency factor, k the Boltzmann constant and T the temperature. By plotting the logarithm of the inverse of the average formation time against the inverse temperature, we obtain a straight line (Fig. 3). From the slope and the intercept of this linear plot, the following data are derived: $E_b = 0.65 \pm 0.05$ eV, and $\nu = 5 \times 10^{12 \pm 1} \text{ s}^{-1}$. For dissociation, E_b is replaced by E_d , the dissociation barrier of the dimer-vacancy

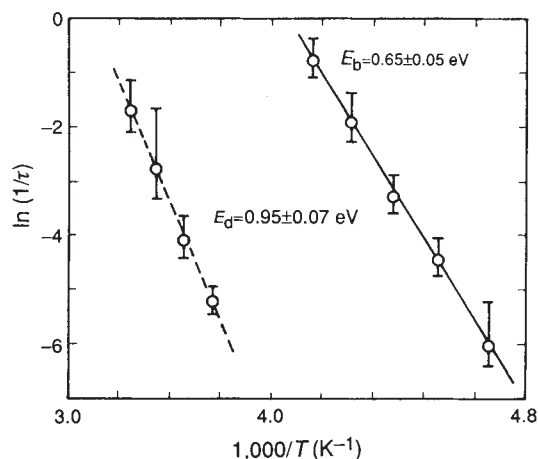


FIG. 4 Formation time and dissociation time of the Re-Ir-vacancy complex as functions of temperature, plotted in the Arrhenius form.

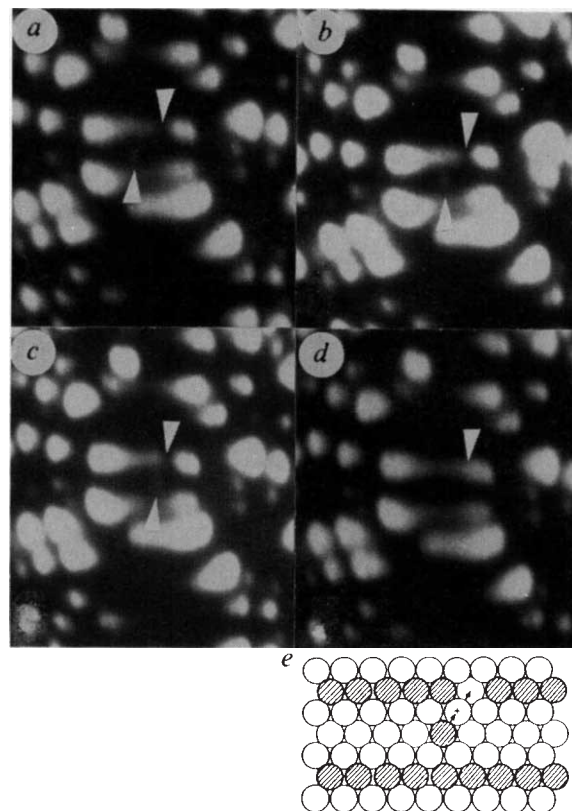


FIG. 5 *a-c*, Random-walk diffusion of a vapour-deposited Ir atom in a double-width (110) surface channel of a partially (1×2) reconstructed Ir (110) surface. One of the channel walls has a lattice vacancy, formed during heating to induce the atomic reconstruction. In *d*, the adatom and the vacancy annihilate one another and this point defect disappears. *e* illustrates this annihilation process. Only shaded atoms are imaged in the FIM.

complex. The second set of data shown in Fig. 3 yields $E_d = 0.95 \pm 0.07$ eV, and $\nu = 5 \times 10^{14 \pm 1} \text{ s}^{-1}$. In this experiment, the detailed atomic steps involved in the atomic replacement are directly seen and the barrier height of each atomic step is measured. Atomic replacement is a possible mechanism for self-diffusion of adatoms, as a random sequence of atomic replacements can occur^{10,11,16}. For a hetero system such as Re on Ir {001}, once the Re adatom has moved into a lattice site, it cannot be restored to an adatom position again by exchanging places with an Ir adatom. Diffusion of the Re adatom ceases. An atomic replacement will incorporate a Re atom into the top Ir surface layer. For heterosystems, atomic replacement is a mechanism for alloy formation of the top surface layer. It will induce surface self-diffusion but not diffusion of the deposited atoms. This process can occur for refractory metals at or below room temperature.

Another atomic process we have observed is adatom-vacancy annihilation. Figure 5 shows a two-row surface channel of a partially (1×2) reconstructed small Ir {110} surface¹⁷. One of the channel walls has a lattice vacancy, formed during the annealing of the surface to induce the atomic reconstruction. To repair this vacancy, an Ir atom is deposited on it as indicated by one of the white arrows. After a random walk by this adatom (Fig. 5*a-c*), the vacancy combines with the deposited adatom, and the point defect disappears (Fig. 5*d*).

Note added in proof: A theoretical calculation¹⁸ shows an intermediate state with an activation barrier of -0.03 eV in the atomic-exchange self-diffusion of Cu {001} surface. In comparison, we measure this barrier height to be 0.95 ± 0.07 eV for Re on an Ir {001} surface. □

Received 19 August; accepted 25 October 1991.

1. Becker, R. S., Golovchenko, J. A. & Swartzentruber, B. S. *Nature* **325**, 419–421 (1987).
2. Eigler, D. M. & Sweizer, E. K. *Nature* **344**, 524–527 (1990).
3. Whitman, L. J., Strosio, J. A., Dragoset, R. A. & Celotta, R. J. *Science* **251**, 1206–1210 (1991).
4. Foster, J. S., Frommer, J. E. & Arnett, P. C. *Nature* **331**, 324–326 (1988).
5. Mamin, H. J., Guethner, P. H. & Ruger, D. *Phys. Rev. Lett.* **65**, 2418–2421 (1990).
6. Lyo, I. W. & Avouris, P. *Science* **245**, 1369–1372 (1989).
7. Bassett, D. W. & Weber, P. R. *Surface Sci.* **70**, 520–528 (1978).
8. Wrigley, J. D. & Ehrlich, G. *Phys. Rev. Lett.* **44**, 661–664 (1980).
9. Kellogg, G. L. & Feibelman, P. J. *Phys. Rev. Lett.* **64**, 3143–3146 (1990).
10. Chen, C. L. & Tsong, T. T. *Phys. Rev. Lett.* **64**, 3147–3150 (1990).
11. Kellogg, G. L. *Phys. Rev. Lett.* **67**, 216–219 (1991).
12. Ehrlich, G. & Stolt, K. *Ann. Rev. Phys. Chem.* **31**, 603–640 (1980).
13. Tsong, T. T. *Rep. Prog. Phys.* **51**, 759–832 (1989).
14. Tsong, T. T. *Atom-Probe Field Ion Microscopy* (Cambridge University Press, 1990).
15. Tsong, T. T. & Chen, C. L. *Phys. Rev.* **B43**, 2007–2017 (1991).
16. Feibelman, P. J. *Phys. Rev. Lett.* **65**, 729–732 (1990).
17. Chen, C. L. & Tsong, T. T. *Phys. Rev. Lett.* **66**, 1610–1613 (1991).
18. Hansen, L., Stolze, P., Jacobsen, K. W. & Nørskov, J. K. *Phys. Rev.* **B44**, 6523–6525 (1991).

ACKNOWLEDGEMENTS. This work was supported by the NSF and in part by the NRC (China) through an award to T.T.T.

Lattice structure of the fullerene ferromagnet TDAE-C₆₀

Peter W. Stephens*, David Cox†, Joseph W. Lauher‡, Laszlo Mihaly*, John B. Wiley§, Pierre-Marc Allemand||, A. Hirsch||, Karoly Holczer§, Q. Li||, Joe D. Thompson¶ & Fred Wudl||

* Department of Physics, State University of New York, Stony Brook, New York 11794, USA

† Department of Physics, Brookhaven National Laboratory, Upton, New York 11973, USA

‡ Department of Chemistry, State University of New York, Stony Brook, New York 11794, USA

§ Departments of Chemistry and Biochemistry and Physics, and Solid State Science Center, University of California, Los Angeles, California 90024, USA

|| Departments of Chemistry and Physics and Institute for Polymers and Organic Solids, University of California, Santa Barbara, California 93106, USA

¶ Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

IN the brief time since the development of techniques for the purification of condensed phases of C₆₀ (refs 1, 2), compounds of C₆₀ have been synthesized that exhibit the important solid-state properties of superconductivity^{3,4} and ferromagnetism⁵. The fact that the C₆₀ molecule participates in these two disparate effects is in itself striking; furthermore, the C₆₀-based material has the highest T_c of any molecular organic ferromagnet⁶. An understanding of the physical origin of this ferromagnetism is sure to require a knowledge of the crystal structure. Here we report on an X-ray diffraction study of the ferromagnet TDAE-C₆₀, where TDAE is tetrakis(dimethylamino)ethylene, C₂N₄(CH₃)₈. We find that the composition is stoichiometric with 1:1 ratio of TDAE to C₆₀. The structure has a *c*-centred monoclinic unit cell suggestive of an anisotropic, low-dimensional band structure. Any theory for the ferromagnetic behaviour will have to take this anisotropy into account.

We prepared samples by allowing an excess of TDAE to react with C₆₀. This differs from previous work⁵ by the absence of any solvent. The sample used here had a low-temperature saturation moment of 0.33 μ_B per formula unit, and a critical temperature T_c = 16.1 K. Figure 1 shows the room-temperature diffraction profile measured at beamline X7A of the National Synchrotron Light Source. The pattern consists of a number of strong, sharp peaks from TDAE-C₆₀, together with several significantly broader peaks from unreacted face-centred-cubic

(f.c.c.) C₆₀ (*a* = 14.18 Å). We have so far been unable to obtain samples free of coexisting unreacted C₆₀, apparently because of the low solubility of that molecule. The powder pattern indexing program TREOR⁷ found only one good candidate unit cell, with figure of merit *F* = 87: this was a *c*-centred monoclinic unit cell, with room-temperature dimensions *a* = 15.874(4) Å, *b* = 12.986(2) Å, *c* = 9.981(3) Å, β = 93.31(1)° (one sigma error bars in last digit). This cell accounted for all of the diffraction peaks other than those assigned to f.c.c. C₆₀. The only systematic absences were found for peaks with *h* + *k* odd, suggesting the space group *C2/m* or the non-centric subgroups *Cm* and *C2*. If the origin is chosen so that the C₆₀ molecules are centred at the 2(*a*) sites (000) and (½½0), the 2(*d*) sites at (½0½) and (0½½) are obvious locations for the TDAE molecules.

The low intensity of diffraction peaks at high angles makes it impossible to refine all of the internal coordinates of the TDAE molecule, and so we assumed a rigid structure for the latter. An ideal molecular geometry was generated using locally developed molecular mechanics programs^{8,9} and is similar to the geometry found for bis(dimethylamino)-methylenemalononitrile¹⁰. The most interesting feature is a predicted 20° twist about the C=C double bond, caused by the excessive steric bulk of the four dimethylamine substituents. Similarly, the C₆₀ molecule was assigned idealized rigid molecule coordinates with bond lengths of 1.37 Å and 1.45 Å, oriented with one of the 2-fold axes along *b*.

The predicted presence of three 2-fold axes and absence of an inversion centre in the TDAE molecule suggest space group *C2* as the logical choice. (Other possibilities such as *C2/m* with disorder in the TDAE orientations must be considered, in analogy to the situation in K₃C₆₀, where disorder in the C₆₀ orientations leads to space group *Fm3m* (ref. 11).) We have performed Rietveld¹² refinements for various possible arrangements of the molecules in the unit cell. A Rietveld fit, excluding the regions of overlap with f.c.c. C₆₀ peaks, is shown in Fig. 1, and the structure is in Fig. 2. The best refinement was given with the long axis of the TDAE molecule (in the direction of the ethylene double bond) assumed to be along the monoclinic *c*-axis. This fit gave temperature factors (assumed isotropic)

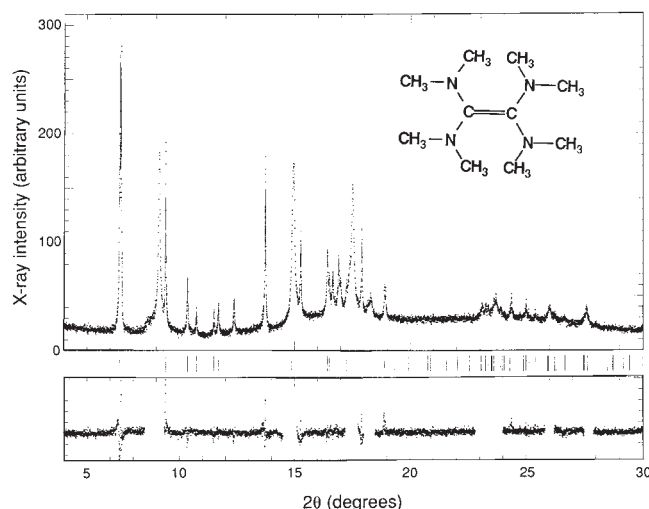


FIG. 1 X-ray diffractogram of TDAE-C₆₀ compound. Data were taken at wavelength 1.2995 Å with a channel-cut Ge(111) monochromator and Ge(220) analyser. The smooth curve is a Rietveld fit, as described in text. To define the baseline, we have drawn a piecewise linear curve through regions of the spectrum where there are no allowed diffraction peaks. Inset shows a diagram of the TDAE molecule. The lower panel shows residuals, to the same vertical scale. Regions around untreated f.c.c. C₆₀ peaks excluded from the fit are absent from the model curve and residuals.

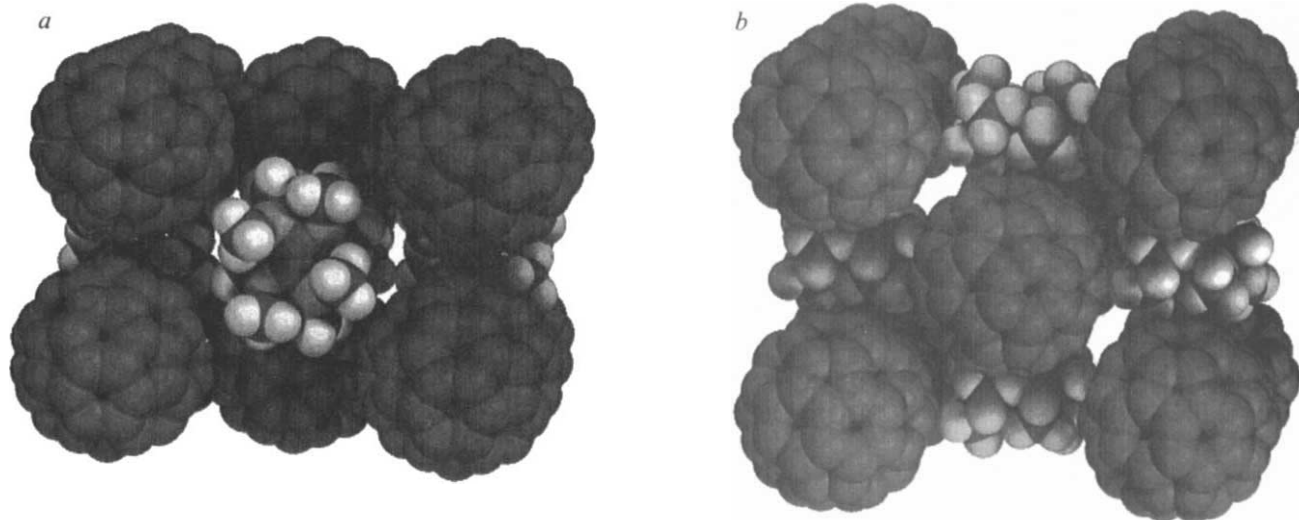


FIG. 2 *a*, Structure of TDAE- C_{60} determined in the present work. View is along the monoclinic *b*-axis. *b*, View along monoclinic *c*-axis.

$B = 0.9 \text{ \AA}^2$ for C_{60} , $3.8 \text{ \AA}^2$ for TDAE, and yielded quality of fit factors $R_1 = 15.7\%$, $R_{WD} = 8.6\%$, $\chi^2 = 1.79$.

We have taken a diffraction pattern of a second sample at 11 K. The degree of contamination by unreacted C_{60} is slightly less in that sample. We achieved similar *R* factors in Rietveld fits. The lattice parameters were $a = 15.807(1) \text{ \AA}$, $b = 12.785(1) \text{ \AA}$, $c = 9.859(1) \text{ \AA}$, $\beta = 94.02(1)^\circ$. We have also studied a sample prepared from C_{60} dissolved in toluene, described in ref. 5, which had a smaller saturation moment of $0.11 \mu_B$ per formula unit. It showed a much smaller amount of coexisting f.c.c. C_{60} , but essentially the same monoclinic pattern. This shows that this is a simple two-phase system, and hence confirms the validity of ignoring the f.c.c. C_{60} peaks in the refinement of Fig. 1. Interestingly, the diffraction lines from this sample were not uniformly sharp; in particular, lines with *h* and *l* both nonzero were so broad as to be virtually undetectable. Evidently, there were solvent molecules present, which acted both to degrade the lattice perfection and the magnetic properties.

Additional refinements showed that the quality of fit is only weakly sensitive to the C_{60} orientation, and that the TDAE occupancy is $100 \pm 10\%$. Other TDAE molecular orientations differing by 45° and 90° rotations are clearly excluded. Difference Fourier maps do not show any significant structure. The nearest C_{60} -TDAE approach is $2.58 \text{ \AA}$, a reasonable distance for a nonbonding C-H distance. Similar refinements in $C2/m$, with the molecules disordered in two sets of mirror-related sites, gave equivalent results.

By crystallographic standards, an *R* factor of 15.7% is not generally regarded as indicating a satisfactory agreement between data and model. It is equally clear that reliable intramolecular positions are not likely to be obtained from such a small number of reflections. But the quality of the data and of the agreement between data and fit is certainly adequate to determine that the unit cell is correct as specified, to exclude greatly different orientations of the TDAE molecule in the unit

cell, and to prove that the composition is 1:1 TDAE to C_{60} . The initial report⁵ that the composition was $TDAE_{0.86}C_{60}$ was based on elemental analysis; in multiphase samples, diffraction experiments are the only accurate means to determine the stoichiometry of the constituent phases. We emphasize that a better refinement will be needed for detailed calculations of band structure.

Our elucidation of the structure helps in the understanding of the ferromagnetism in this material. It was argued previously that TDAE- C_{60} is an itinerant ferromagnet⁵: that is, the unpaired spins are carried by delocalized electrons. We have measured the electron paramagnetic resonance signal in this material, and find a relatively low value for the *g*-factor, $g = 2.0008$. This implies that the open-shell species reside principally on the C_{60} , and to a lesser extent on the TDAE.

Now that we have determined the unit cell parameters, we see that indeed the C_{60} cores have intermolecular contacts of the same order as observed for the metallic alkali-doped materials. At room temperature, the intermolecular separation is $9.98 \text{ \AA}$ along the *c*-axis, and $10.25 \text{ \AA}$ within the *a*-*b* plane. In comparison, in the metallic alkali-doped samples, the observed distances^{11,13} range from 10.07 to $10.25 \text{ \AA}$. In a tight-binding picture, the electronic bandwidth is proportional to the overlap of electronic wave functions centred on adjacent fullerenes, which decreases exponentially with separation. The observation in this material of the short interfullerene distance, along with the much greater $10.25\text{-\AA}$ distance within the *a*-*b* plane, leads us to conclude that the TDAE salt should have a very anisotropic band structure. This stands in contrast to the superconducting A_3C_{60} structures, which are all cubic^{11,13}. It is well known that low-dimensional bands are subject to a variety of electronic and magnetic instabilities that lead to solid-state phase transitions^{14,15}. We therefore conclude that in TDAE- C_{60} , the ferromagnetically ordered state cannot be understood without considering the band anisotropy. □

Received 30 October; accepted 24 December 1991.

- Krätschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347**, 354-358 (1990).
- Ajje, H. et al. *J. phys. Chem.* **94**, 8630-8633 (1990).
- Hebard, A. et al. *Nature* **350**, 600-601 (1991).
- Holczer, K. et al. *Science* **252**, 1154-1157 (1991).
- Allemand, P. M. et al. *Science* **253**, 301-303 (1991).
- Kinoshita, M. et al. *Chem. Lett.* 1225-1228 (1991).
- Werner, P.-E., Eriksson, L. & Westdahl, M. *J. appl. Cryst.* **18**, 367-370 (1985).
- Lauher, J. W. *J. Am. chem. Soc.* **108**, 1521-1531 (1986).
- Lauher, J. W. *J. molec. Graphics* **8**, 34-38 (1990).

- Adhikesavalu, D. & Venkatesan, K. *Acta crystallogr.* **C39**, 589-592 (1983).
- Stephens, P. W. et al. *Nature* **351**, 632-634 (1991).
- Rietveld, H. M. *J. appl. Cryst.* **2**, 65-71 (1969).
- Fleming, R. M. et al. *Nature* **352**, 787-788 (1991).
- Emery, V. J. in *Chemistry and Physics of One Dimensional Metals*, ed. Keller, H. J. (Plenum, New York, 1977).
- Jacobs, I. S. et al. *Phys. Rev.* **B14**, 3036-3051 (1976).

ACKNOWLEDGEMENTS. Work at Brookhaven was supported by the U.S. Department of Energy, Division of Materials Sciences and at Stony Brook and Santa Barbara by the NSF. Work at Los Alamos was done under the auspices of the U.S. Department of Energy.

Energetics of negatively curved graphitic carbon

Thomas Lenosky*, Xavier Gonze*, Michael Teter*†
& Veit Elser*

* Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA

† Applied Process Research, SP-PR-22, Corning Incorporated, Corning, New York 14831, USA

By analogy with the positively curved carbon networks that comprise the fullerenes¹⁻³, it has been suggested⁴ that negative curvature might be possible in graphitic carbon sheets, giving rise to extended structures corresponding to periodic minimal surfaces⁵ that divide space into two disjoint labyrinths. Whereas the positive curvature of fullerenes results from the presence of five-membered rings, negative curvature would derive from seven-membered rings. Here we present calculations of the cohesive energy and bulk moduli of two such hypothetical, negatively curved carbon networks. We find that both have a cohesive energy smaller than that of graphite but significantly greater than that of C₆₀, even though the proportion of odd-membered rings is comparable. We therefore suggest that it is worth scrutinizing the insoluble residue generated in the carbon-arc preparation of fullerenes⁶ for possible evidence of fragments of negatively curved graphitic carbon.

The fullerenes owe their stability to their similarity with graphite. For small curvatures of a graphitic sheet, carbon atoms can maintain their preferred *sp*² bonding while allowing the sheet to have various three-dimensional geometries. In the fullerenes, curvature is introduced in discrete amounts by the addition of five-membered rings, exactly twelve being necessary to form a closed spheroid. A seemingly plausible alternative⁴ is to introduce negative curvature by means of larger rings, in particular, rings with seven and eight carbon atoms. If we specifically consider an *sp*²-bonded carbon network without boundary having *n*₅ five-membered rings, *n*₇ seven-membered rings and an indeterminate number of six-membered rings, then Euler's theorem gives the constraint

$$n_5 - n_7 = 6\chi \quad (1)$$

where χ , the Euler characteristic, is a topological index of the corresponding surface of pentagons, hexagons and heptagons. Local rearrangements of bonds in an *sp*² network can simultaneously change *n*₅ and *n*₇ while keeping their difference constant, as required by topology. Clearly removing both five- and seven-membered rings by mutual annihilation is energetically desirable, but this process cannot rid the network of all defect rings unless $\chi = 0$. If the surface is viewed topologically as a sphere with *g* handles, then $\chi = 2 - 2g$. Thus a complex topology (many handles) requires a large excess of seven-membered rings.

We have constructed what we believe to be the most symmetrical *sp*² bonded structures having acceptable curvature and only six- and seven-membered rings. Our structures are periodic in three dimensions and have $\chi = -4$ per unit cell. Equation (1) then requires that each unit cell have exactly 24 seven-membered rings. A structure having only the 24 seven-membered rings per unit cell (56 atoms) has very nonplanar bonding and would be the negative curvature analogue of the 20-atom dodecahedron having only five-membered rings. As in C₆₀, the planarity of the bonds is improved by adding six-membered rings: 20 in the case of C₆₀, 80 per unit cell for each of the structures here, giving 216 atoms in total. The bond networks are shown in Fig. 1. In many ways the correspondence with C₆₀ is close. In particular, the seven-membered (defect) rings are always completely surrounded by six-membered rings and are equivalent by symmetry. Both structures have cubic symmetry and bear a strong resemblance to periodic minimal surfaces. The structure shown in Fig. 1a can be viewed as a tetrahedral confluence of four

TABLE 1 Physical properties of four cubic forms of carbon

Structure	ρ^*	$\Delta E^\dagger$	$B^\ddagger$
schwarzite D	1.15	0.18	9.4
schwarzite P	1.02	0.20	7.5
f.c.c. C ₆₀	1.71 (ref. 10)	0.42 (ref. 7)	1.4 (ref. 10)
diamond	3.52	0.02 (ref. 11)	44.3

* Density in g cm⁻³.

† Total energy relative to graphite in eV per atom.

‡ Bulk modulus in 10¹¹ dyne cm⁻².

pores, repeated in space to form a diamond structure. Figure 1a shows one such confluence (or half the unit cell) viewed along a [100] direction of the conventional cubic unit cell. The other structure, Fig. 1b, corresponds to a confluence of six pores which is repeated in space to form a simple cubic structure. In this case there is one confluence per unit cell. The view along a [110] direction is shown. Following Mackay and Terrones⁴, we will refer to our structures using the designations of the minimal surfaces they resemble. Thus the structures shown in Fig. 1a and b correspond to the D (sometimes called F) and P surfaces respectively. We propose that the class of structures considered here be given the name 'schwarzite', partly in anticipation of their likely colour but mainly to honour the mathematician H. A. Schwarz who first explored the triply periodic minimal surfaces⁵. Here the minimal surface idea is at best descriptive. In fact, close examination of Fig. 1a (and b) shows that the analogy is defective in one important respect: the carbon net does not divide space into equivalent labyrinths as is well known for the D (and P) minimal surface (for further discussion, see below).

For calculating the geometries and energies of the two forms of schwarzite, we used a potential energy function appropriate to a nearly planar bond network. Contributions to the total energy can be ranked by order of magnitude as follows: (1) bond stretching, (2) bond bending, (3) π -electron resonance and (4) interplanar van der Waals interaction. We treat the first of these essentially as a rigid constraint, thus imposing a fixed length *r*₀ on all the bonds in the structure. That still leaves one and a half degrees of freedom per atom to be fixed by contributions (2) to (4). Bond-bending forces are sufficient to determine the structure completely (graphite being an exception) so that (4) can safely be neglected. Finally, we believe that an approximate treatment of (3) which includes only resonance between pairs of adjacent, possibly misaligned, *p_z* orbitals is adequate here. The potential function used in our calculations contains three unknown bond-bending parameters ϵ_1 , ϵ_2 , ϵ_3 :

$$\Delta E = \epsilon_0 \sum_{(ij)} \frac{1}{2} (r_{ij} - r_0)^2 + \epsilon_1 \sum_i (\mathbf{u}_{i(1)} + \mathbf{u}_{i(2)} + \mathbf{u}_{i(3)})^2 + \epsilon_2 \sum_{(ij)} (1 - \mathbf{n}_i \cdot \mathbf{n}_j) + \epsilon_3 \sum_{(ij)} (\mathbf{n}_i \cdot \mathbf{u}_{ij})(\mathbf{n}_j \cdot \mathbf{u}_{ij}) \quad (2)$$

The first term above constrains the bond length. With this term alone, the network is highly underconstrained so that the exact value of ϵ_0 (provided it is large) is irrelevant. With *r*₀ = 1.42 Å, the bond length in graphite, ΔE gives the energy difference with respect to a single graphite sheet. For C₆₀, as well as both of the schwarzite structures studied, the contribution of the bond-stretching term to ΔE was less than 1%. The expressions for the three bond-bending terms involve unit vectors $\mathbf{u}_{ij}$, which point from carbon atom *i* to a neighbour *j*, as well as the unit vectors $\mathbf{n}_i$, which are normal to the plane determined by the three neighbours of carbon atom *i* (designated *i*(1), *i*(2), *i*(3) above). We view the first bond-bending term as arising from perturbations of the electronic structure near individual atoms, and the remaining two terms are our attempt to approximate the energetics of π -electron resonance. Two additional terms of the atomic type (having independent quadratic variations away from graphite and expressible as bi- and triquadratic combinations

of the unit vectors $\mathbf{u}$) might have been included as well as many more terms involving the unit vectors $\mathbf{n}$. Our choice is the simplest we could think of that would at least be able to distinguish between splay and torsional misalignments of p_z orbitals, possibly an important difference in the energetics of positively and negatively curved carbon surfaces.

The bond-bending parameters ε_1 , ε_2 and ε_3 were determined by fits to local density approximation (LDA) total energy calculations of variously perturbed configurations of a $\sqrt{7} \times \sqrt{7}$ unit cell of graphite (14 atoms). A total of five sets of small, out-of-plane displacements were used to fit the four independent transverse force constants of the structure. Our LDA calculations⁷ used the conjugate gradient algorithm⁸ with a carbon pseudopotential adjusted to give accurate total energies with a

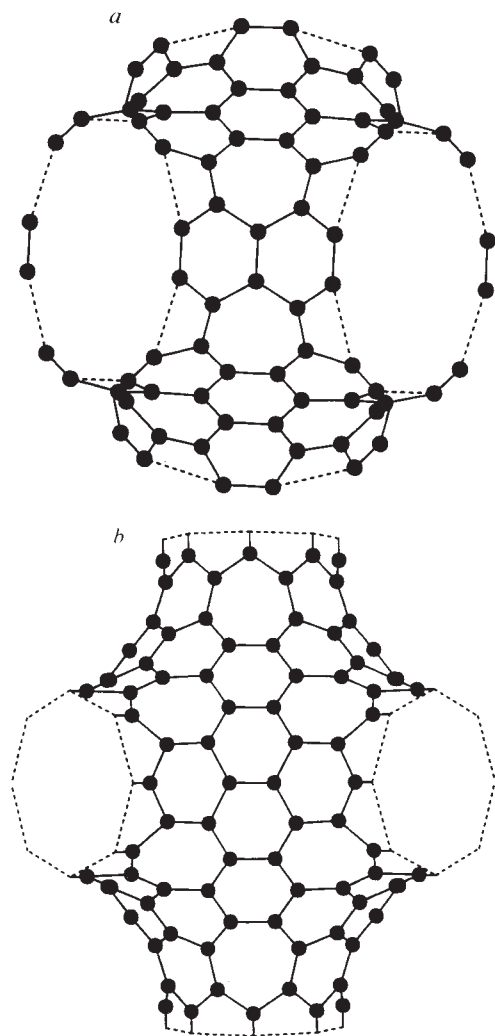


FIG. 1 Two-fold symmetry axis views of the cubic D (a) and P (b) forms of schwarzite, emphasizing differences in the layout of the seven-membered rings. Both images show two sets of three seven-membered rings which (in three dimensions) symmetrically surround a planar, 3-fold symmetric, hexagon. The centres of these symmetric hexagons correspond to points where a minimal surface would have zero curvature. There are four such points in *a* forming a regular tetrahedron and eight in *b* at the vertices of a cube (only two visible in each case). Spheres circumscribing the latter polyhedra define our cavity diameter d_{c1} . The carbon net in *a* is terminated where it passes through the faces of a regular tetrahedron. The 12-sided polygons thus formed contain the diameters of six carbon hexagons. The centres of these hexagons themselves form a regular hexagon whose diameter defines our pore diameter d_{p1} . In *b*, six pores are cut by the faces of a cube forming octagons having 4-fold symmetry. The edge centres of one of these octagons describe another octagon whose vertices fall on one circle with diameter d_{p1} .

TABLE 2 Characteristic dimensions of schwarzite

Structure	a	d_{c1}	d_{p1}
D	24.7	9.9	7.4
P	16.2	11.7	5.4

Lattice parameter a (of the conventional unit cell) and smaller cavity and pore diameters, d_{c1} and d_{p1} (see Fig. 1), of the D and P forms of schwarzite. All dimensions are in Å.

plane-wave basis set having a 15-Hartree (~ 408 -eV) cutoff. Isolated graphite monolayers were simulated by spacing periodic graphite sheets from 5 to 8 Å. Sets of 4, 8 and 32 special points were used to sample the Brillouin zone⁹. Our results, $(\varepsilon_1, \varepsilon_2, \varepsilon_3) = (0.96, 1.29, 0.05)$ eV, indicate that splay and torsional misalignments of the normal vectors give about the same reduction in π -electron resonance. The validity of the three-parameter model was checked by an LDA calculation of the energy of a C_{60} molecule having equal bond lengths. This calculation, using first one and then eight special points in the Brillouin zone⁹, gave an energy difference $\Delta E = 0.38(3)$ eV per atom with respect to a graphite monolayer. The three-parameter model predicts 0.52 eV per atom for the same quantity. An SCF calculation¹⁰ of the relaxed molecule gives 0.42 eV/atom⁷. The atomic positions, lattice parameter a and energy ΔE of each schwarzite structure were calculated by minimizing the potential energy function, equation (2). Bulk moduli were calculated by fitting the quadratic change in ΔE with respect to lattice parameter.

Our calculations give both cubic forms of schwarzite a cohesive energy that is ~ 0.2 eV per atom greater than C_{60} (Table 1). According to our decomposition of the bond-bending energy into atomic and resonance contributions, this result can be traced to the geometrical fact that a seven-membered ring can have 120° bond angles whereas a five-membered ring cannot. Thus in C_{60} the quantity $(\mathbf{u}_{ii(1)} + \mathbf{u}_{ii(2)} + \mathbf{u}_{ii(3)})^2$ has the value 0.38 for each atom, whereas its largest value in the schwarzites is only 0.05 (D) and 0.06 (P), the average values being even smaller: 0.02 (D) and 0.03 (P). If one measures p_z orbital misalignments in terms of an angle θ defined by $\cos \theta = \mathbf{n}_i \cdot \mathbf{n}_j$ (using the definition of $\mathbf{n}$ above), then for some bonds in the schwarzites θ is as large as 35° , whereas both bonds in C_{60} have $\theta \approx 23^\circ$. The average of $\mathbf{n}_i \cdot \mathbf{n}_j$ is however slightly larger for both of the schwarzites. The bulk modulus of the schwarzites is understandably much smaller than that of diamond, as the lattice constant is determined by bond-bending forces.

The geometrical characteristics of these very open structures can be quantified by a set of pore and cavity diameters in addition to the lattice constant a . The pore diameters have a minimum where they pass through the faces of a regular tetrahedron, in the case of the D structure (Fig. 1a), or cube, in the case of the P structure (Fig. 1b). Precise definitions of these diameters, d_{p1} , are given in Fig. 1. At the confluence of pores, a cavity is formed which contains a centre of cubic symmetry. We define the diameter of the cavity, d_{c1} , to be the diameter of the sphere that just touches the centres of the four (in D) and eight (in P) hexagons having perfect 3-fold symmetry. Corresponding diameters of the pores and cavities of the complementary labyrinth are given by the relations $\sqrt{2}(d_{p1} + d_{p2}) = \sqrt{4/3}(d_{c1} + d_{c2}) = a$ for the D structure, and $d_{p1} + d_{p2} = (d_{c1} + d_{c2})/\sqrt{3} = a$ for the P structure. From the values in Table 2 we see that both schwarzite structures deviate from minimal surfaces because $d_{p1} \neq d_{p2}$ and $d_{c1} \neq d_{c2}$. Finally we note that even the smallest pores have diameters rivaling those of the known zeolites¹¹, while the larger cavity in the P structure ($d_{c2} = 16.4$ Å) can comfortably accommodate a C_{60} molecule.

We have repeated our calculations for the 192-atom P-surface network of Mackay and Terrones⁴ and find: density $\rho = 1.16$, $\Delta E = 0.19$, bulk modulus $B = 13.3$, $a = 14.9$ (units as in Tables

1 and 2). Although the curvature in this structure is twice as concentrated in the form of eight-membered rings, it seems to be energetically just as viable.

Negatively curved graphitic structures correspond to a very deep metastable state of carbon, the two cubic schwarzites being in fact preferred energetically over the smallest fullerene, C_{60} . Although our calculations have focused on only two of the most symmetrical schwarzites, we believe that our general conclusions apply to random pore geometries as well. The occurrence in nature of these metastable states is a different problem altogether. If graphitic carbon grows by accretion, then fullerenes and schwarzites correspond respectively to extreme modes of slow and fast growth. During slow growth, the minimization of dangling bond energy in a quasi-equilibrium low-temperature environment clearly favours positive curvature and has been advanced as the mechanism for fullerene formation¹². On the other hand, the number of growth sites (dangling bonds) on a negatively curved surface grows more rapidly than even in planar (single sheet) graphite. Thus the opposite extreme, namely explosive crystallization in a highly supersaturated vapour, might be the most favourable environment for the formation of schwarzite. Because a range of conditions probably exists in the carbon-arc fullerene synthesis of Krätschmer *et al.*⁶, we consider

it worthwhile to investigate in more detail the insoluble by-products. A different approach to schwarzite synthesis (D. Muller, personal communication) is to fuse together fullerenes in a film by intense electron or ion bombardment. The density of schwarzite is somewhat lower than that of fullerene films but may be increased, as our calculations suggest, if fullerenes are trapped in the schwarzite cavities. □

Received 23 September; accepted 26 November 1991.

1. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
2. Kroto, H. W. *Science* **242**, 1139–1145 (1988).
3. Curl, R. F. & Smalley, R. E. *Science* **242**, 1017 (1988).
4. Mackay, A. L. & Terrones, H. *Nature* **352**, 762 (1991).
5. Schwarz, H. A. *Gesammelte mathematische Abhandlungen* (Springer, Berlin, 1890).
6. Krätschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347**, 354–358 (1990).
7. Pickett, W. E. *Comp. Phys. Rep.* **3**, 115–197 (1989).
8. Teter, M. P., Payne, M. C. & Allan, D. C. *Phys. Rev.* **B40**, 12255–12263 (1989).
9. Monkhorst, H. J. & Pack, J. D. *Phys. Rev.* **B13**, 5188–5192 (1976).
10. Schulman, J. M. & Disch, R. L. *J. chem. Soc. Chem. Commun.* **6**, 411–412 (1991).
11. Estermann, M., McCusker, L. B., Baerlocher, C., Merrouche, A. & Kessler, H. *Nature* **352**, 320–322 (1991).
12. Kroto, H. W. & McKay, K. *Nature* **331**, 328–331 (1988).
13. Fischer, J. E. *et al. Science* **252**, 1288–1290 (1991).
14. Brewer, L. *Lawrence Berkeley Laboratory Rep.* No. LBL-3720.

ACKNOWLEDGEMENTS. Funding for this work was provided by the NSF (T.L.) and the David and Lucille Packard Foundation (V.E.).

Optical modification of a single impurity molecule in a solid

Th. Basché & W. E. Moerner*

IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, California 95120-6099, USA

THE possibility of obtaining information about solids on a truly microscopic scale has stimulated several recent advances in the optical detection and spectroscopy of single impurity centres in solids. For the system composed of pentacene impurity molecules in the crystal *p*-terphenyl, absorption¹ and fluorescence excitation² studies at liquid-helium temperatures have led to direct observations of the lifetime-limited homogeneous linewidth of a single pentacene molecule³, as well as the surprising observation of spontaneous spectral diffusion in a crystal⁴. Spectral diffusion—changes in the resonance frequency of an impurity molecule with time as a result of structural relaxation processes in the molecular environment—is generally expected in amorphous hosts. We report here the observation of optical spectra of single perylene impurity molecules in a polymeric (polyethylene) host. At 1.5 K, individual perylene molecules show the expected spectral diffusion; furthermore, we observe light-induced changes in resonance frequency (that is, persistent spectral hole-burning⁵) for certain single molecules. These observations suggest the possibility of optical storage at the single-molecule level.

Samples were prepared by predissolving zone-refined perylene in acetone, mixing with low-density poly(ethylene) (PE) powder ($\leq 25\%$ crystallinity), drying at 323 K for 30 min under high vacuum, pressing films at 413–423 K between glass slides, and quenching to 77 K for 1 min. The resulting samples, 10–20 μm thick, had excellent optical clarity (low scattering), little or no aggregate formation as tested by low-resolution optical absorption, and perylene/PE mass ratios of 9.5×10^{-7} . The samples were mounted in a low-temperature optical cryostat at the joint focus of a lens providing the excitation beam and a parabolic mirror collecting the emitted fluorescence, as described in detail elsewhere⁶. The tunable excitation laser (linewidth ~ 2 –3 MHz) generated wavelengths near the centre

and to the red of the inhomogeneous line of the perylene (0–0) $S_1 \leftarrow S_0$ transition in PE (442–450 nm). The emitted fluorescence was long-pass filtered with a low-fluorescence filter to reject Rayleigh scattered radiation. Fluorescence excitation spectra were obtained by scanning the dye laser and recording the red-shifted fluorescence with a GaAs phototube and a photon counter.

By tuning the laser to the long-wavelength side of the inhomogeneous line, we could easily observe isolated single-molecule spectra, as in trace *a* of Fig. 1. Moreover, even with low laser-probing intensity, various linewidths are observed at 1.5 K, as shown in traces *a*–*f*. The observed linewidths range from 52 ± 2 MHz in trace *a* to 142 ± 5 MHz in trace *f*, and all of these values are larger than the lifetime-limited width of 24.9 MHz (ref. 7). The different linewidths are not correlated with the position relative to the centre of the inhomogeneous line or with the probing intensity. In fact, trace *a* of Fig. 1 was recorded with the highest probing power of 9 nW (in a ~ 5 - μm -diameter laser focal spot). This suggests that spectral diffusion on the 1–100-MHz scale is occurring during the ~ 2 s required

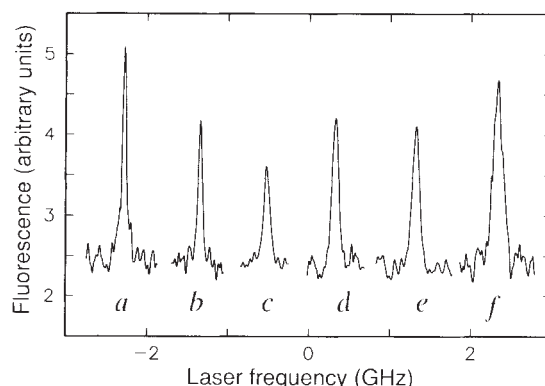


FIG. 1 Fluorescence excitation spectra of single molecules of perylene in PE at 1.5 K. The excitation wavelengths and power levels, respectively, are (a) 448.438 nm, 9 nW; (b) 448.021 nm, 9 nW; (c) 450.152 nm, 4.7 nW; (d) 448.020 nm, 5 nW; (e) 447.875 nm, 4.5 nW; (f) 448.452 nm, 9 nW. Photon count interval is 100 ms, detection bandwidth is 10 Hz. The different signal amplitudes occur because of the different power levels and because some molecules may not be exactly in the centre of the laser spot.

* To whom correspondence should be addressed.

to measure the lineshape. In addition, there may be a pure homogeneous dephasing (nondiagonal) contribution to the width at this temperature due to fast configurational changes of the polymer host present at 1.5 K. According to the standard models⁸, this contribution to the width is generally assumed to be the same for all molecules. In addition to the range of linewidths, we also observe slower spectral diffusion which appears as discontinuous jumps in resonance frequency on the scale of several hundred megahertz between or during the 25-s laser scans, in a fashion similar to previous work on pentacene in *p*-terphenyl⁴.

The spectral diffusion process in amorphous solids has generally been treated as follows. The low-energy excitations characteristic of the amorphous state are approximated by a collection of asymmetric double-well potential energy surfaces in the configurational coordinates of the host^{9,10}. At low temperatures, only the two lowest energy levels of each double-well are important, and these excitations are therefore referred to as two-level systems (TLSs). Tunnelling through the barrier separating the two wells of an asymmetric TLS often involves emission or absorption of a low-energy phonon to compensate for the intrinsic energy mismatch. Such phonon-assisted tunnelling processes are thought to be partially responsible for both spectral diffusion and dephasing in amorphous solids through the coupling between the impurity centre and the TLSs^{11,12}. Both diagonal perturbations, which shift the energy of the optical transition of the impurity, and off-diagonal perturbations, which shift the phase of the excited state, are expected¹³. Because of the unusual nature of the observed spectral jumps for our single molecules, a detailed analysis of the spectral diffusion must be the subject of future work. Here we concentrate on the novel light-driven process for single molecules, persistent spectral hole-burning.

Usually, persistent spectral hole-burning in solids refers to optical modification of an inhomogeneous line where the number of molecular centres in resonance is much greater than one. When light absorption induces photochemical changes in the excited centres or photophysical (nonphotochemical) alterations in the nearby host, a narrow spectral 'hole' or dip can

develop in the absorption spectrum because the altered centres no longer absorb at the laser wavelength. (By definition, 'persistent' requires that the alteration persists longer than any excited state lifetime, and in practice, hole lifetimes from seconds to many years have been demonstrated.) In the single-molecule regime, however, only the isolated homogeneous absorption profile of a single centre is present. When photochemical or photophysical changes occur as a result of optical excitation, the resonance frequency of the single molecule may move far away from the laser scan range, that is, the absorption line seems to vanish.

Figure 2 shows an example of persistent spectral hole-burning for a single perylene molecule in PE. This molecule was deliberately chosen to have little spontaneous spectral jumping from one resonance frequency to another. Before the traces shown in the figure, the molecule was scanned five times and no significant changes were observed. Traces *a*, *b* and *c* show three additional scans of the molecule. After trace *c*, the laser was tuned into resonance with the molecule using the scanning power of 9 nW; within 30 s, the size of the emitted fluorescence suddenly dropped to the background level, indicating hole-burning. Trace *d* was then acquired, which shows that the resonance frequency shifted by more than ± 1.25 GHz as a result of the light-induced change. The exact location of the new resonance frequency is unknown at present; by analogy with previous nonphotochemical hole-burning in large ensembles of molecules¹⁴, however, the shift may be expected to be up to 100 cm^{-1} . A further scan some minutes later (trace *e*) showed that the molecule surprisingly returned to resonance at the original wavelength, within 10 MHz. Further scans (traces *f* and *g*) revealed no further changes in resonance frequency. After trace *g*, the laser was again tuned into resonance until the fluorescence dropped; a final scan (trace *h*) showed that the molecule was again absent from the range of the laser scan. (We note that other single molecules were observed to burn irreversibly: that is, after burning the molecules did not return to the original wavelength during observation times of up to 20 min. The behaviour reported in Fig. 2 was, however, more common.)

To provide more evidence that the hole-burning process shown in Fig. 2 is laser-driven and not very slow coincidental spectral diffusion, we studied the kinetics of the process. Figure 3*a* and *b* shows time scans with a fixed laser frequency. In trace *a*, the laser was tuned into resonance with the molecule in the first 10 s at a power level of 4.5 nW. The eventual abrupt drop of the fluorescence signal at 260 s is the hole-burning event; the return of the emitted fluorescence indicates that the molecule has returned to its original frequency, only to be burned again, and so on. In trace *b* of the figure, the laser power level was increased to 27 nW; now the time in resonance, although stochastic, is clearly shorter. This indicates that the average rate of persistent spectral hole-burning for this perylene impurity in PE increases with laser power. (Separate measurement¹⁵ of many burning events for a single molecule at a fixed power level shows that the distribution of burning times is exponential.) In Fig. 3*c*, the average burn times for several single molecules as obtained from time scans like traces *a* and *b* are plotted against the laser power used. Even though the actual burn times are stochastic, the average burn time for all centres investigated decreases with increasing laser power. In contrast, and as expected, there is no correlation between the average return times and the laser power (Fig. 3*d*).

The generally accepted model for nonphotochemical hole-burning of impurity molecules in amorphous hosts^{16,17} involves transitions among TLSs near the impurity as a result of the optical excitation. The exact means by which this occurs are unknown, but one model¹⁶ assumes that the barrier height to configurational changes of the host molecules is lower when the impurity is in the optical excited state. On the other hand, spontaneous (phonon-assisted) transitions of TLSs far from the impurity affect the dephasing or optical linewidth. Because a

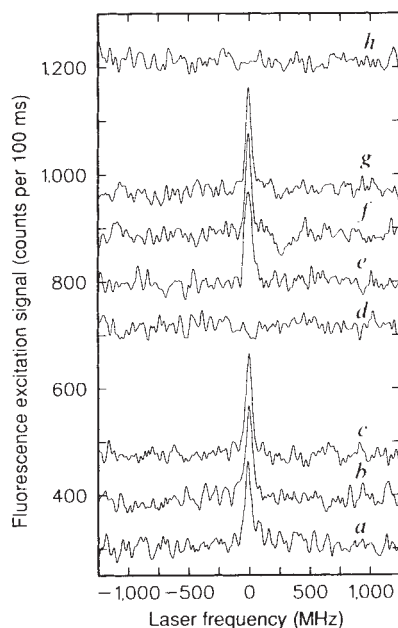


FIG. 2 Persistent spectral hole-burning of a single perylene molecule in PE at 1.5 K. The content of scans *a*–*h* is described in the text. Zero detuning corresponds to a laser wavelength of 448.021 nm, laser power for scanning and burning is 9 nW, detection bandwidth is 10 Hz and scan time is 25 s.

transition of a TLS near the impurity should produce a large change in the local strain field, the resulting shift in the optical resonance frequency should be large, whereas frequency shifts from distant TLSs produce small frequency shifts. Our data provide direct evidence for the following general picture of a hierarchy of TLSs surrounding the impurity. The TLSs far away produce fast shifts which are small in magnitude; the closer TLSs produce larger shifts which are more infrequent; and TLSs very close to the impurity are effectively locked until persistent spectral hole burning occurs. After the hole-burning event, the return of the optically excited centre to exactly the same wavelength in Fig. 2 is surprising, as it suggests that coupling to exactly one nearby TLS is involved in the hole-burning process in this system. If this is true, the single impurity molecule may in the future serve as a direct probe of a single TLS.

The ability to modify optically the absorption of single impurity centres in a solid leads naturally to the possibility of optical storage at the single-molecule level. One can imagine a very thin layer of a material with a very broad inhomogeneous line so that single molecules are isolated and spread over a large range of frequency space. The resonance frequencies constitute the addresses of all the bits to be encoded in a single focal volume. A binary sequence of '1's and '0's can be produced by altering or ignoring each single molecule absorption. Of course,

to achieve areal density higher than available in current hole-burning schemes, a method of producing optical beams much smaller than the diffraction limit must be used. This concept is as yet highly speculative, but among its advantages are the extreme areal density (up to perhaps 10^{14} bits cm^{-2}) and the greatly increased signal from each molecule resulting from the use of sub-diffraction-limited beams (the signal-to-noise ratio scales roughly as $\sqrt{\sigma P/A}$, where σ is the peak absorption cross-section, P is the laser power and A is the beam area). Some disadvantages are the required low temperatures, stochastically variable burning times and variable frequency locations of the individual bits from laser spot to laser spot. Nevertheless, single-molecule persistent spectral hole-burning not only provides a window into the photophysics and low-temperature dynamics of the amorphous state, but also allows such novel optical storage schemes to be contemplated. $\square$

Received 10 October; accepted 11 November 1991.

1. Moerner, W. E. & Kador, L. *Phys. Rev. Lett.* **62**, 2535–2538 (1989).
2. Orrit, M. & Bernard, J. *Phys. Rev. Lett.* **65**, 2716–2719 (1990).
3. Moerner, W. E. & Ambrose, W. P. *Phys. Rev. Lett.* **66**, 1376 (1991).
4. Ambrose, W. P. & Moerner, W. E. *Nature* **349**, 225–227 (1991).
5. *Persistent Spectral Hole-Burning: Science and Applications* (ed. Moerner, W. E.) *Topics in Current Physics*, Vol. 44 (Springer, Berlin, 1988).
6. Ambrose, W. P., Basché, Th. & Moerner, W. E. *J. chem. Phys.* **95**, 7150–7163 (1991).
7. Birks, J. B. *Photophysics of Aromatic Molecules*, 123 (Wiley-Interscience, London, 1970).
8. Silbey, R. & Kassner, K. *J. Lumin.* **36**, 283–292 (1987).
9. Phillips, W. A. (ed.) *Amorphous Solids: Low Temperature Properties* (Springer, Berlin 1981).
10. Anderson, P. W., Halperin, B. I. & Varma, C. M. *Phil. Mag.* **25**, 1–9 (1972).
11. Small, G. J. in *Spectroscopy and Excitation Dynamics in Condensed Molecular Systems* (eds Agranovich, V. M. & Hochstrasser, R. M.) 515–554 (North-Holland, Amsterdam, 1983).
12. Friedrich, J. & Haarer, D. in *Optical Spectroscopy of Glasses* (ed. Zschokke, I.) 149–198 (D. Reidel, 1986).
13. Hayes, J. M., Jankowiak, R. & Small, G. J. *Persistent Spectral Hole-Burning: Science and Applications* (ed. Moerner, W. E.) *Topics in Current Physics*, Vol. 44, Ch. 5 (Springer, Berlin, 1988).
14. Lee, I.-J., Hayes, J. M. & Small, G. J. *J. chem. Phys.* **91**, 3463–3469 (1991).
15. Basché, Th., Ambrose, W. P. & Moerner, W. E. *J. opt. Soc. Am. B* (submitted).
16. Hayes, J. M., Stout, R. P. & Small, G. J. *J. chem. Phys.* **74**, 4266–4275 (1981).
17. Jankowiak, R. & Small, G. J. *Science* **237**, 618–625 (1987).

ACKNOWLEDGEMENTS. This work was supported in part by the US Office of Naval Research.

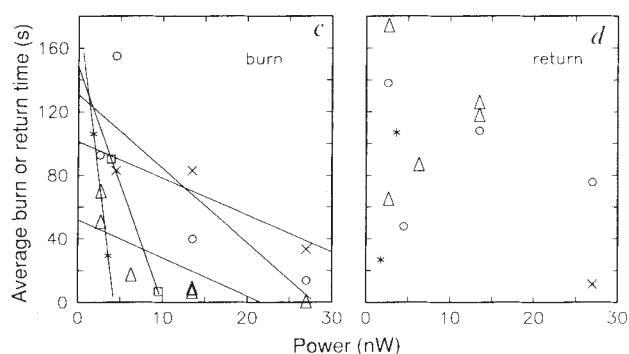
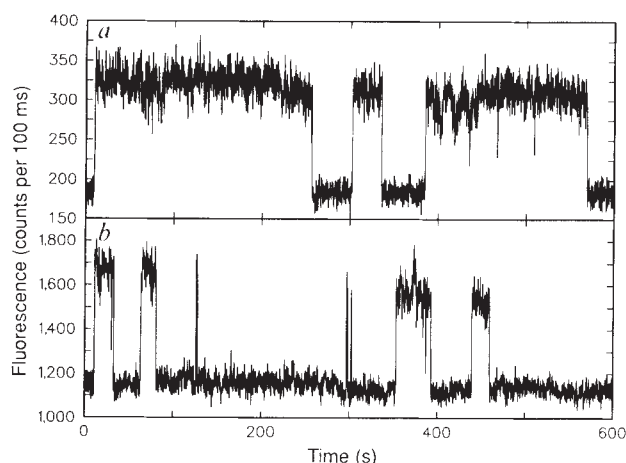


FIG. 3 Power dependence of the hole-burning process. *a*, Time scan at a fixed laser wavelength of 448.159 nm with 4.5 nW laser power. *b*, Time scan at the same wavelength with 27 nW laser power. *c*, Average burn times against laser power, with each point derived from 2–8 burning events. The different symbols represent five different single molecules, and the lines are simple linear fits to the data for each molecule to guide the eye only. *d*, Average return times against laser power for four of the single molecules shown in (*c*). The number of data points at the highest power is limited because irreversible persistent spectral hole burning was more common in this case.

A molecular water pump in quartz dislocations

Malcolm I. Heggie

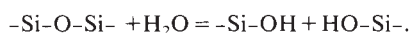
Scientific Computing Group, Department of Computer Science, Prince of Wales Road, University of Exeter, Exeter EX4 4PT, UK

RECENTLY Bakker and Janssen¹ have reported experiments suggesting that water leaks out of bubbles in quartz preferentially along dislocations against thermodynamic gradients in fugacity, chemical potential and pressure. This effect indicates that the density of gas-rich fluid inclusions cannot be calculated from metamorphic pressure–temperature conditions and, conversely, that the density of the fluid cannot be used to infer metamorphic pressure–temperature conditions. Using computer modelling techniques, I show here that water not only diffuses easily along dislocations in quartz, but can also be ‘pumped’ along them. The simulations indicate that in dislocation cores, water molecules dissociate and the protons and hydroxyl groups become strongly bound to kinks. A shear stress can do work on kinks to move them along a dislocation, dragging with them these water-bearing species and sweeping other, undissociated water molecules before them.

The computer simulation uses a simple interatomic potential for silica^{2,3}, based on the Keating potential⁴, with an extra oxygen–oxygen interaction applied to clusters of up to 6,000 atoms and unit cells of 500 atoms. The interaction between the quartz lattice and the diffusing molecule neglects hydrogen bonding (which is weak for compressed water molecules⁵), and I assume that diffusion of water in channels occurs without

dissociation. The same approach has been used to describe the enhancing effect of water on the mobility of dislocations⁶, and hence a mechanism for the hydrolytic weakening of quartz. Models of the most significant dislocations in quartz were built, and the structure and movement of one of the best-characterized dislocations, a basal 60° dislocation (Burgers vector $\mathbf{b} = \frac{1}{3}[\bar{1}2\bar{1}0] \equiv \mathbf{a}_2$, $\mathbf{l} = \frac{1}{3}[2\bar{1}\bar{1}0] \equiv \mathbf{a}_1$, labelled B60a), was then studied in detail. I considered a mechanism for dislocation motion in wet quartz that allows a considerable reduction in kink-pair formation energy compared with dry quartz. The reduction (from 5 to 1 eV) occurs when a bond in the core of the dislocation can be hydrolysed in the manner suggested by Griggs *et al.*⁷ and the resulting hydroxyl groups can separate from each other during kink migration.

Two properties of quartz combine to reduce this phenomenon to an essentially topological one. The first is the high bond energy of the Si-O bond (~8 eV), making consideration of the SiO₂ network paramount; the second is the system of helical channels inherent to the α -quartz network. Locally the bonding can be disrupted by a water molecule by the hydrolysis reaction:



Ab initio calculations of total energy⁸ indicate that this reaction will be spontaneous for a bond stretched beyond 4% of its normal α -quartz length, for which the energy cost is ~0.1 eV.

The channels in right-handed α -quartz (which I shall consider here) are lined with either a single or a double helix of Si-O-Si linkages, and channels sharing a linkage have opposite handedness. Along the direction of a basal lattice vector, two adjacent linkage-sharing channels have no net 'helicity', h , because a left-handed single helix ($h = -1$) neighbours a right-handed single helix ($h = 1$). Any two such adjacent channels can have their row of common linkages sheared and then rebonded to give two systems of rings ($h = 0$). This is effectively a screw dislocation dipole, and it would cost ~3–4 eV per lattice vector to create one in perfect α -quartz. In the core of a dislocation, however, there are fewer constraints on topology, and the two channels of the core can much more easily interconvert between rings and helices. For B60a (Fig. 1 is an axial view) there is already a screw component to the Burgers vector which adds

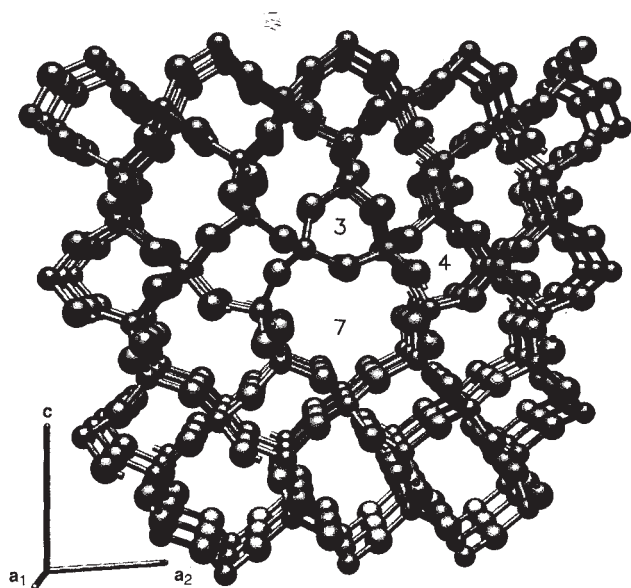


FIG. 1 Axial perspective of the dislocation 'B60a' in the $\frac{1}{3}[\bar{1}2\bar{1}0]$ {0001} system along $[2\bar{1}\bar{1}0]$; the larger atoms are oxygen. The three lattice vectors of the hexagonal lattice, $\mathbf{c}$ (up the page), $\mathbf{a}_1$ (the dislocation axis) and $\mathbf{a}_2$ (the Burgers vector), are shown bottom left. The triangular (3) channel and the channel to its right (4) are picked out in Fig. 2. The heptagonal (7) channel is the favoured location for undissociated water molecules.

TABLE 1 Insertion and migration energies (eV) in basal plane

Structure	Channel	Insertion energy	Migration energy	Activation energy
perfect (H4H4)	hexagon	1.20	0.58	1.78
perfect (H4H4)	rectangle	1.34	0.52	1.86
dislocation (C3C4)	heptagon	0.45	0.56	1.01
dislocation (C3H4)	heptagon	0.33	0.90	1.23
dislocation (H3C4)	heptagon	0.50	0.43	0.93
dislocation (H3H4)	heptagon	1.58	0.19	1.74

–1 to the net helicity of the three channels in the core. The three core channels are lined, respectively, with 3, 4 and 7 Si atoms in projection, arising from perfect channels of 4, 4 and 6 Si atoms. If we label channels according to the number of Si atoms in projection and precede this with a label describing helicity (H, right-handed helix; C, circles or rings; H', left-handed helix) then the fourfold channels in perfect quartz are H4 and the six fold channels H'6. The screw dislocation dipole referred to earlier would be labelled C4C6. In B60a the four lowest-energy 'reconstructions' allow the three- and fourfold channels to be either rings or right-handed helices, and they are therefore labelled C3C4, C3H4, H3C4 and H3H4. They are illustrated in Fig. 2, without the sevenfold channel below them, which has a helicity such that the sum over the three channels is –1 (complete labelling would be C3C4C7, C3H4H'7, H3C4H'7 and H3H4H'7, with H' standing for a left-handed double helix).

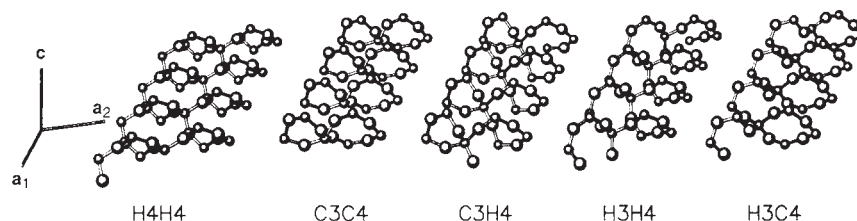
Energetically, the last three 'reconstructions' are practically degenerate, whereas the first is much higher in energy. Dislocation unit cells of 500 atoms, periodic along the line direction but with free boundary conditions, have energies 6.59, 5.08, 5.07 and 5.18 eV, respectively, for C3C4, C3H4, H3C4 and H3H4.

The potential energy surface for a dislocation in a simple solid is corrugated, with the valleys (Peierls valleys) separated by full lattice vectors. Alternative reconstructions give rise to intervening valleys. Calculations⁹ show that dislocations in quartz move between these different reconstructions during glide. In particular, the ground state of H3C4 can jump to H3H4 before jumping to its ground state in the next Peierls valley. This facilitates glide by reducing the 'height' of the kinks that allow the dislocation to jump from one valley to the next. Practically, this process can only happen in wet quartz, because every jump that converts a C channel to H must involve a dangling bond where the helix terminates, which can only be stable if protonated or hydroxylated. This gives rise to one source of pumping of water along dislocations in quartz: water-bearing species (protons and hydroxyl groups arising from dissociated water molecules) are bound to kinks and dragged along with them as the kinks are moved by an applied shear stress. There is, however, an extra factor to be taken into account: the role of undissociated water molecules in the lattice.

I have studied the passage of water molecules through channels in the basal plane of quartz (hexagonal and rectangular, referring to channels bounded by six and four silicon atoms in projection, respectively) and through dislocation channels in the basal plane. The oxygen atom of the water molecule was stepped along each channel in the $\mathbf{a}_1$ direction in 0.1-Å steps and allowed to relax in all but the $\mathbf{a}_1$ direction, as were the 5,796 atoms in the cluster, which comprised 12 layers of 483-atom cylinders with fixed boundaries.

The results in Table 1 demonstrate that the enlarged heptagonal channel of the dislocation is extremely favourable for the water molecule. The total activation energy for diffusion is about the same (1.8 eV) for basal diffusion along either type of channel in perfect α -quartz, but strongly reduced in the dislocation cores of three of the four reconstructions for B60a. There are detailed geometrical considerations that affect the insertion and migration energies of water along these dislocation cores,

FIG. 2 The four low-energy states of the core of the basal dislocation B60a. There are four reconstructions labelled C3C4, C3H4, H3H4 and H3C4, in which either the 3 or 4 channel can be a system of rings (C) or right-handed helices (H). For reference the lattice vectors and the equivalent channels (H4H4) in perfect α -quartz are shown on the left.



but it also seems that there is an overriding topological one. The largest free volume for the water molecule arises when the heptagonal channel is in either the ring (C3C4) or single helix form (C3H4, H3C4). The double helix formed in H3H4 is somewhat stretched and radially contracted, pushing the insertion energy up.

What is clear from these calculations is that although the different reconstructions of B60a are almost degenerate, the presence of water could cause C3H4 and H3C4 to be the preferred structures. The low energies of insertion found here indicate that there is almost enough free volume for uncompressed water molecules in these reconstructions. Hydrogen bonding and entropy may well reduce the insertion free energy practically to zero. Some enigmatic results on annealing wet quartz samples suggest that water simply disappears from bubbles (S. H. Kirby and A. K. Kronenberg, personal communication) and perhaps a reasonable explanation of this phenomenon is the restructuring of dislocation cores to accommodate water during annealing.

The final and most interesting consequence is that water contained in dislocation cores in H3C4 or C3H4 reconstructions can be pumped by an applied stress. Confined as they are to the heptagonal channels of H3C4 or C3H4 (insertion energy 0.3–0.5 eV), they will not enter nearby perfect channels (insertion energy 1.2–1.4 eV). During dislocation motion a segment of H3H4 will be nucleated on the dislocation and will repel the water molecule (insertion energy 1.6 eV). As this segment expands under the influence of an applied shear stress, water molecules will be repelled and pushed before it. The energetics are illustrated in Fig. 3 where a segment of two lattice vectors of H3H4 has been nucleated in a 12-layer, 5,796-atom cluster of H3C4, and a water molecule has been stepped through it. It is clear that H3H4 acts like a piston, as predicted from Table 1.

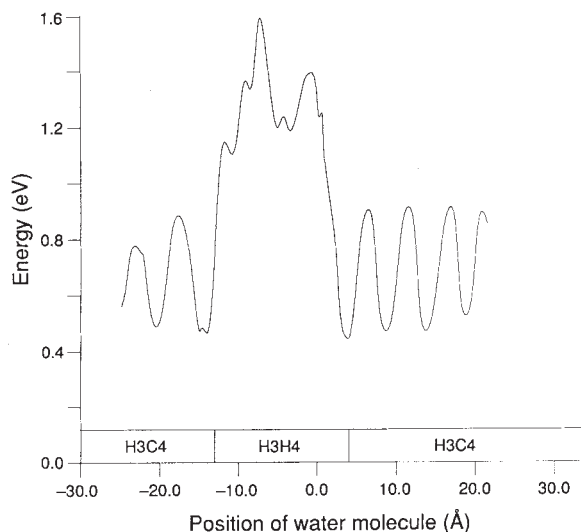


FIG. 3 The energy of a water molecule in a 5,796-atom cluster of SiO_2 around an H3C4 dislocation with a small (two-lattice-vector) segment of H3H4, as a function of the position of a water molecule along the heptagonal channel. Note the blocking effect of the H3H4 segment and the kinks that delimit it.

On energy grounds, single-kink nucleation is normally easier than kink-pair nucleation. Where a section of H3C4 dislocation penetrates a bubble, it could fill with water, and an applied stress (or differential thermal expansion caused by annealing) could nucleate a kink and a segment of H3H4. This would push the water into the bulk (possibly against other thermodynamic gradients) away from the bubble. And it could do so cyclically, a second kink returning the dislocation to H3C4 character. This process also underlines the importance of microcracking in the early experiments on the hydrolytic weakening of α -quartz^{7,9}. The stress enhancement at crack tips creates and moves wet dislocations which are fed by a reservoir of water in the crack. The dislocations themselves facilitate the spread of water into the crystal and in so doing catalyse their own motion. □

Received 29 August; accepted 21 November 1991.

1. Bakker, R. J. & Janssen, J. B. H. *Nature* **345**, 58–60 (1990).
2. Heggie, M. I. & Nylén, M. *Phil. Mag.* **B50**, 543–555 (1984).
3. Heggie, M. I. & Nylén, M. *Phil. Mag.* **B51**, L69–72 (1985).
4. Keating, P. N. *Phys. Rev.* **145**, 637–645 (1966).
5. Heggie, M. I., Jones, R., Latham, C. D., Maynard, S. C. P. & Tole, P. *Phil. Mag. B* (in the press).
6. Heggie, M. I. & Jones, R. *Phil. Mag.* **A53**, L65–70 (1986).
7. Griggs, D. T. & Blacic, J. D. *Science* **147**, 292–295 (1965).
8. Heggie, M. I. & Jones, R. *Phil. Mag.* **55**, L47–51 (1987).
9. Kronenberg, A. K., Kirby, S. H., Aines, R. D. & Rossman, G. R. *J. geophys. Res.* **91**, 12723–12741 (1986).

ACKNOWLEDGEMENTS. This work was supported by the SERC.

Indirect chemical effects of methane on climate warming

Jos Lelieveld & Paul J. Crutzen

Max-Planck-Institute for Chemistry, PO Box 3060, D-6500 Mainz, Germany

METHANE concentrations in the atmosphere have increased from about 0.75 to 1.7 p.p.m.v. since pre-industrial times^{1,2}. The current annual rate of increase of about 0.8% yr⁻¹ (ref. 2) is due to increases in industrial and agricultural emissions. This increase in atmospheric methane concentrations not only influences the climate directly, but also indirectly through chemical reactions. Here we show that the climate effects of methane's atmospheric chemistry have previously been overestimated, notably by the Intergovernmental Panel on Climate Change (IPCC)³, largely owing to neglect of the height dependence of certain atmospheric radiative processes. Using available estimates of fossil-fuel-related leaks of methane, our results show that switching from coal and oil to natural gas as an energy source would reduce climate warming. A significant fraction of methane emissions cannot, however, be accounted for by known sources; should leakages from gas production and distribution be underestimated for some countries, then it might be unwise to switch to using natural gas.

A concentration increase of an atmospheric gas that absorbs infrared radiation initially decreases the net flux of long-wave terrestrial radiation F (W m^{-2}) to space. The Earth-atmosphere system responds to this with warming at the surface. The climate forcing per unit increase of a gas i is denoted here by f_i . We have calculated the current climate forcing of CH_4 relative to

that of CO_2 , $\text{RF}_{\text{CH}_4} = f_{\text{CH}_4}/f_{\text{CO}_2} = 26$ (mol per mol), adopting the infrared band model of Ramanathan^{4,5}. Results obtained with this band model agree closely both with measurements and with detailed infrared radiative-transfer calculations⁵. Our RF_{CH_4} value is higher than that presented by IPCC³, $\text{RF}_{\text{CH}_4} = 21$, which is based on a different band model⁶.

The relative forcing RF_{CH_4} indicates the instantaneous relative climate forcings of CH_4 and CO_2 . But as CO_2 and CH_4 have very different lifetimes in the atmosphere, a more appropriate ratio for several applications is the (relative) global warming potential^{3,7-9}, which indicates the time-integrated climate forcings of these gases over a selected period T

$$\text{GWP} = \frac{\int_0^T f_{\text{CH}_4} R_{\text{CH}_4}(t) dt}{\int_0^T f_{\text{CO}_2} R_{\text{CO}_2}(t) dt} \quad (1)$$

where $R_i(t)$ describes the atmospheric decay of a pulse of CH_4 or CO_2 . In its simplest representation this has the form

$$R_i(t) = e^{-t/\tau_i} \quad (2)$$

where τ_i is the atmospheric lifetime of gas i . Recent studies indicate that $\tau_{\text{CH}_4} \approx 10$ years^{10,11}. On the other hand, τ_{CO_2} is not well quantified, depending on a multitude of physical, chemical and biological processes, both terrestrial and marine. A more comprehensive description of $R_{\text{CO}_2}(t)$ involves multiple exponential decay terms, each accounting for uptake in a biogeochemical reservoir (for example, the ocean surface layer or deep ocean), with its own time constant for CO_2 removal. According to the IPCC³, which adopted the formulation by Siegenthaler¹², the effective CO_2 lifetime is ~ 120 years for a time horizon of about a century. GWPs calculated over long time horizons, however, have large uncertainties⁹.

A factor which complicates calculations of the climate forcing of methane is that growing CH_4 emissions also affect atmospheric chemistry, leading to concentration increases of the greenhouse gases O_3 , H_2O and CO_2 (refs 3, 7-9). An increase of CH_4 in NO_x -rich environments leads to more ozone in the troposphere and the lower stratosphere¹³ ($\text{NO}_x = \text{NO} + \text{NO}_2$). The oxidation of CH_4 is also an important source of water vapour in the stratosphere¹⁴. Methane is ultimately oxidized to CO_2 . But biogenic CH_4 is recycled CO_2 , so that only the fossil-fuel-related CH_4 emissions need to be considered. Thus, equation (1) should formally be extended into

$$\text{GWP}^* = \frac{\int_0^T [(f_{\text{CH}_4} + \sum f_i (\partial X_i / \partial X_{\text{CH}_4}) R_{\text{CH}_4}(t) + G_{\text{CO}_2}(t))] dt}{\int_0^T f_{\text{CO}_2} R_{\text{CO}_2}(t) dt} \quad (3)$$

in which f_i is the climate forcing by O_3 or stratospheric H_2O , due to a change of their mixing ratio X_i caused by an increase of X_{CH_4} . The term $G_{\text{CO}_2}(t)$ accounts for the conversion of fossil-fuel-derived CH_4 to CO_2

$$G_{\text{CO}_2}(t) = -\frac{\partial}{\partial t} R_{\text{CH}_4}(t) \int_t^T f_{\text{CO}_2} R_{\text{CO}_2}(t') dt' \quad (4)$$

We calculated GWP^* by applying a radiative-convective one-dimensional coupled climate-chemistry model to both hemispheres, to account for different ocean covers and trace gas emissions¹⁵. Interhemispheric exchange of long-lived species and heat fluxes into the oceans are included. The model resolves 86 levels up to 65 km altitude, and adopts a fixed cloud cover and height, and constant relative humidity at the Earth's surface. It considers spectral overlap among CO_2 , CH_4 , N_2O , O_3 and H_2O . The decay of CO_2 follows a function $R_{\text{CO}_2}(t)$ (ref. 15) that relates CO_2 concentration changes to its emissions during the industrial age¹⁶⁻¹⁸. Table 1 presents our own and the IPCC³ estimates of GWP^* , with the conventional assumption of constant 1990 gas concentrations. Our estimates of the indirect chemical effects on the climate are much less than those of IPCC. We believe that the feedbacks through tropospheric O_3

TABLE 1 The GWP and GWP^* of methane for different time horizons

Emission scheme	T (yr)	GWP	GWP^*	Due to O_3	Due to H_2O	Due to $\text{CO}_2(F)^\dagger$	CH_4 -OH feedback
Constant concentrations (this work)	10	18.4	24.9	5.4	1.1	—	—
	20	13.6	18.7	4.2	0.9	0.1	—
	30	10.6	14.4	3.1	0.7	0.2	—
	50	7.4	9.6	1.7	0.5	0.3	—
	100	4.3	5.4	0.8	0.3	0.5	—
Constant concentrations (IPCC ³)	20	9.6	22.9	8.7	3.6	1	—
	100	2.2	7.6	2.9	1.5	1	—
Increasing emissions scenario B	10	18.4	26.9	5.5	1.1	—	1.9
	20	13.6	21.7	4.4	0.9	0.1	2.8
	30	10.6	18.2	3.5	0.7	0.2	3.4
	50	7.4	13.0	2.0	0.5	0.3	3.1
	100	4.3	7.5	0.9	0.3	0.5	2.0

The GWP and GWP^* are calculated in p.p.b.v./p.p.b.v. The asterisk to GWP^* denotes that indirect chemical effects are taken into account. To convert the units from p.p.b.v./p.p.b.v. into mass/mass, the numbers given should be multiplied by the ratio of the CO_2 and CH_4 relative molecular masses, 2.75. T is the time horizon. 'Constant' refers to calculations in which emissions are assumed to remain constant in the future, with τ_{CH_4} fixed. 'Increasing' refers to calculations based on IPCC³ emission scenario B. The columns following GWP^* specify the indirect chemical effects. A dash means not considered or negligible.

[†] Only applies to fossil-fuel-related CH_4 , a distinction that was not made in IPCC³.

and stratospheric H_2O have been strongly overestimated by IPCC, mainly because of their neglect of the altitude dependence of the O_3 and H_2O climate forcings. Increased photochemical O_3 formation in the troposphere is particularly important in environments containing catalytically active NO_x . Because NO_x , which has an atmospheric lifetime of only a few days¹³, is produced predominantly from fossil fuel combustion, it is most abundant in the lower troposphere of the industrial regions of the Northern Hemisphere. Ozone additions to the lower troposphere are, however, much less effective in climate forcing than additions to the upper troposphere and lower stratosphere¹⁹, so that calculations should not be based on the increase in total tropospheric ozone, as was done by IPCC. A similar argument applies to stratospheric H_2O production. Although CH_4 oxidation is an important source of stratospheric water vapour, its total amount, especially in the lower stratosphere which is the level most relevant to climate, is dominated by transport from the troposphere, that is, the flow through the cold trap of the tropical tropopause²⁰. Although CH_4 oxidation roughly doubles H_2O levels in the upper stratosphere, this contributes only slightly to climate forcing.

Another significant factor in defining GWP^* is that the reaction between CH_4 and OH is a sink for both gases, so that the increase of CH_4 can cause a depletion of OH and thus an increase of τ_{CH_4} with time²¹. Because of the many chemical interactions between gases, determination of the GWP^* including CH_4 -OH feedback requires a description of all trace gas emissions that affect OH concentrations. This implies that for most of the troposphere CH_4 , CO and NO_x should be considered. Whereas increases in CH_4 and CO lead to lower OH concentrations, additions of NO_x have the opposite effect, because of its catalytic role in the formation of tropospheric O_3 , the precursor of OH radicals, and the reaction $\text{NO} + \text{HO}_2 \rightarrow \text{NO}_2 + \text{OH}$. Actually, because of the short lifetime of NO_x , quantifying all chemical feedbacks requires the use of global chemistry-transport models with high spatial resolution. As clouds affect tropospheric photochemistry, leading to less O_3 , NO_x and OH, a proper description of aqueous-phase chemical processes is also needed²². Models that account for all of these factors have not yet been developed, but preliminary results have been obtained with simplified models. Two-dimensional models indicate that as much as 30% of the currently observed CH_4 increase may result from a reduction in OH concentrations, and 70% from increased emissions^{11,23}. For an increase of

0.8% yr⁻¹ in CH₄, this implies that global average OH concentrations may be decreasing by ~0.25% yr⁻¹.

To show the dependence of GWP* on the chosen scheme, Table 1 also lists values obtained by adding equal pulses of CO₂ and CH₄ to IPCC³ emission scenario B, in which it is assumed that CO₂ and CH₄ continue to increase throughout the next century. We corrected this IPCC emission scheme with respect to CFCs, which will be phased out by the year 2000. For the scheme based on scenario B, we obtain GWP* values ~10% and 40% higher than our constant-concentration GWP* for $T = 10$ and $T = 100$ years, respectively. Our scenario-B-based GWP* is higher than our direct GWP by 45% and 75% for $T = 10$ and $T = 100$ years. Fortuitously, our scenario-B-based GWP* values are almost equal to those given by IPCC for the fixed concentration scenario³ (Table 1). In addition to the above-mentioned differences, the reduction in CFC emission that we adopted for scenario B implies less O₃ destruction in the stratosphere, so that less ultraviolet radiation reaches the troposphere. Consequently, OH formation is reduced, thus increasing τ_{CH_4} . Further effects arise from overlaps of the infrared absorption bands of CO₂, CH₄ and O₃ with those of H₂O, because of the expected increase of atmospheric water vapour related to global warming. Clearly, assessment of indirect effects of atmosphere chemistry on GWP* depends on the model and the emission scheme adopted.

Finally, we address the question of how the climate forcing caused by emissions of CO₂ and CH₄ responds to shifts between coal, oil and natural gas as energy sources. First, we must briefly consider fossil-fuel-related CH₄ emissions. The current total CH₄ source^{10,24} is $520 \pm 80 \text{ Tg yr}^{-1}$ (1 Tg = 10^{12} g). Important biogenic sources are wetlands, rice paddies and ruminants²⁵. In addition, a large fraction of the CH₄ emissions, about $110 \pm 20 \text{ Tg yr}^{-1}$, is ¹⁴C-free²⁶. This 'old' methane escapes through leaks from natural gas production and distribution systems, from oil production and coal mines, from methane hydrate deposits and melting permafrost regions. Destabilization of CH₄ hydrates is probably a minor source of ~5 Tg yr⁻¹ (ref. 25). Wetlands and rice fields seem to emit only small amounts of 'old' methane²⁷. Methane emissions from coal mining are estimated at $35 \pm 10 \text{ Tg yr}^{-1}$ (ref. 25). A few studies^{28,29} indicated that worldwide release of CH₄ as oil-associated gas is $14 \pm 6 \text{ Tg yr}^{-1}$. Extrapolation to the entire world of losses estimated for natural gas systems in several countries which are members of the Organization for Economic Cooperation and Development (OECD)²⁸⁻³⁰ yields 5–10 Tg yr⁻¹. If correct, this would leave a missing 'old' CH₄ source of $50 \pm 35 \text{ Tg yr}^{-1}$. It has been speculated that large gas losses occur in the Soviet Union, where 40% of the world gas production takes place. But more work on releases of 'old' CH₄, not only from fossil-fuel-related but also from biological sources is needed, including rice fields in which old organic matter may have been ploughed up.

Following Rodhe^{8,31}, we calculated the fractional leakage e_g from natural gas use that, if exceeded, would increase climate warming after switching from coal and oil to natural gas. Note that e_g (also e_c for coal and e_o for oil) is the specific CH₄ emission in mol per mol carbon produced or, assuming 100% burning efficiency, per mol CO₂ emitted. A switch between coal and gas, for constant total energy production, does not have climate consequences if

$$\Delta C + \Delta G + \text{GWP}^*(e_c \Delta C + e_g \Delta G) = 0 \quad (5)$$

C and G are the emissions of CO₂ by coal and gas combustion. With j_c , j_o and j_g the specific energy yields for coal, oil and gas combustion (MJ per mol CO₂ emitted)

$$j_c \Delta C + j_g \Delta G = 0 \quad (6)$$

Consequently,

$$\text{GWP}^* \left(e_g - e_c \frac{j_g}{j_c} \right) - \frac{j_g}{j_c} + 1 = 0 \quad (7)$$

TABLE 2 Break-even points

T (yr)	GWP* (mol/mol)	$e_g(\text{coal})$ (%)	$e_g(\text{oil})$ (%)
10	26.9	4.3–5.7	2.4–2.9
20	21.8	4.9–6.3	2.8–3.3
30	18.4	5.5–7.0	3.2–3.7
50	13.3	7.0–8.4	4.2–4.7
100	8.0	10.5–12.0	6.7–7.2

Estimated fractional leakages associated with the use of natural gas (the break-even points) at which the net gain from switching energy production from coal and oil to natural gas ($e_g(\text{coal})$ and $e_g(\text{oil})$ respectively) is nullified. Note that the GWP* based on emission scenario B includes the contribution of $G_{\text{CO}_2}(t)$ (equation 3) for fossil-fuel-related CH₄ (see Table 1 under CO₂(F)). Ranges express uncertainties in CH₄ releases associated with coal and oil production.

From ref. 29 we adopt a value of $j_g/j_c = 1.64$. We rearrange and correct for the implicit assumption that natural gas consists of 100% CH₄, whereas 90% is more realistic⁸. For comparisons of gas with coal and oil, we now distinguish between $e_g(\text{coal})$ and $e_g(\text{oil})$, respectively. Hence

$$e_g(\text{coal}) = \frac{0.64(\text{GWP}^*)^{-1} + 1.64e_c}{0.9} \quad (8)$$

Similarly, for a switch between oil and gas, adopting $j_g/j_o = 1.44$ (ref. 29), we derive

$$e_g(\text{oil}) = \frac{0.44(\text{GWP}^*)^{-1} + 1.44e_o}{0.9} \quad (9)$$

On the basis of the coal- and oil-related CH₄ emissions mentioned above^{25,28,29} and CO₂ emission estimates from ref. 32, we obtain $e_c = 1.3 \pm 0.4\%$ for coal and $e_o = 0.5 \pm 0.15\%$ for oil, yielding the ranges for $e_g(\text{coal})$ and $e_g(\text{oil})$ in Table 2. Our results indicate that, with $T = 10$ years, the fractional natural gas leakage should be less than 2.4–2.9% to attain a reduction in climate forcing by switching from oil to gas; gas would be preferable to coal for a gas leakage less than 4.3–5.7%. For longer GWP* time horizons, natural gas comes out more favourably. As noted, studies in several OECD nations have indicated that gas losses are well below 1%, implying substantial reductions in climate forcing following shifts from oil and particularly from coal to natural gas. But gas-loss estimates from other parts of the world (such as the Soviet Union) are not known. Considering the 'old' CH₄ budget deficit of $50 \pm 35 \text{ Tg yr}^{-1}$, a fractional leakage of less than 1% may not occur everywhere. Therefore, the consequences of switching from coal and oil to natural gas must be considered for each country, debiting climate effects not only to the producers, but also to the consumers. □

Received 15 July; accepted 29 October 1991.

- Pearman, G. I. *et al. Nature* **320**, 248–250 (1986).
- Steele, L. P. *et al. J. Atmos. Chem.* **5**, 125–171 (1987).
- Intergovernmental Panel on Climate Change (IPCC) *Report of Working Group I* (eds Houghton, J. T. *et al.*) (WMO/UNEP, New York, 1990).
- Ramanathan, V. *J. Atmos. Sci.* **33**, 1330–1346 (1976).
- Donner, L. & Ramanathan, V. *J. Atmos. Sci.* **37**, 119–124 (1980).
- Hansen, J. *et al. J. Geophys. Res.* **93**, 9341–9364 (1988).
- Derwent, R. G. *Trace Gases and their Relative Contribution to the Greenhouse Effect, Rep. AERE-R13716* (Harwell Laboratory, Oxfordshire, 1989).
- Rodhe, H. *Science* **248**, 1217–1219 (1990).
- Lashof, D. A. & Ahuja, D. R. *Nature* **344**, 529–531 (1990).
- Vaghjiani, G. L. & Ravishankara, A. R. *Nature* **350**, 406–409 (1991).
- Valentin, K. M. thesis, Univ. of Mainz (1991).
- Siegenthaler, U. *J. Geophys. Res.* **88**, 3599–3608 (1983).
- Crutzen, P. J. *Pure appl. Geophys.* **106-8**, 1385–1399 (1973).
- Nicolet, M. *Disc. Faraday Soc.* **37**, 7–27 (1964).
- Brühl, C. & Crutzen, P. J. *Clim. Dynam.* **2**, 173–203 (1988).
- Neftel, A. *et al. Nature* **295**, 220–223 (1982).
- Raynaud, D. & Barnola, J. M. *Nature* **315**, 309–311 (1985).
- Keeling, C. D. *et al. in Carbon Dioxide Review 1982* (ed. Clark, W. C.) 377–398 (Clarendon, Oxford, 1982).
- Wang, W.-C. *et al. J. Atmos. Sci.* **37**, 545–552 (1980).
- Dobson, G. M. B., Brewer, A. W. & Cwilog, B. M. *Proc. R. Soc. Lond.* **A185**, 144–175 (1946).

21. Sze, N. D. *Science* **195**, 673–675 (1977).
22. Lelieveld, J. & Crutzen, P. J. *Nature* **343**, 227–233 (1990).
23. Isaksen, I. S. A. & Hov, O. *Tellus* **B39**, 271–285 (1987).
24. Prinn, R. et al. *J. geophys. Res.* (submitted).
25. Cicerone, R. J. & Oremland, R. S. *Global biogeochem. Cycles* **2**, 299–327 (1988).
26. Wahlen, M. et al. *Science* **245**, 28C–290 (1989).
27. Quay, P. D. et al. *Global biogeochem. Cycles* **2**, 385–397 (1988).
28. Arthur, D. Little (ADL) *Methane Emissions from the Oil and Gas Production Industries, Rep. to Ruhrgas A. G.* (Essen, 1989).
29. Okken, P. A. *Energy Policy*, 203–204 (March, 1990).
30. *Ermittlung der Methan-Freisetzung durch Stoffverluste bei der Erdgasversorgung der BRD* (Battelle, Frankfurt, 1989).
31. Lelieveld, J., Crutzen, P. J. & Brühl, C. *Chemosphere* (submitted).
32. Marland, G. & Rotty, R. M. *Tellus* **B36**, 232–261 (1984).

ACKNOWLEDGEMENTS. We thank C. Brühl for his help with the computations.

TABLE 1 Sample depth below unconformity, X and $\delta^{13}\text{C}$ of $\text{Fe}(\text{CO}_3)\text{OH}$ component in Neda Formation goethite

MHD#	Sample	Depth (cm)	X	$\delta^{13}\text{C}$ (‰)*
1126	OPWis-2 (ultrasonic fines)	0–5	0.0029	–14.6
1123	OPWis-4	12–16	0.0033	–15.2
1178	90-OPWis-G-1	17–18	0.0042	–16.0
1098	OPWis-5	28–31	0.0055	–16.9
1133	OPWis-6	50–57	0.0048	–17.0
1090	OPWis-9	102–107	0.0054	–17.4
1131	IRMWis-1	150–153	0.00645	–17.6

$\delta^{13}\text{C} = [R(\text{sample})/R(\text{standard}) - 1]$; $R = {}^{13}\text{C}/{}^{12}\text{C}$; standard is PDB.

Formation goethite that explains all of these observations.

The variations of X and $\delta^{13}\text{C}$ with depth (Fig. 1) are generally consistent with a system in which CO_2 undergoes steady-state diffusive transport through the 'soil' to the Earth's atmosphere^{3,11,12}. Higher soil CO_2 partial pressures and more negative $\delta^{13}\text{C}$ values (Fig. 1) arise from *in situ* oxidation of organic matter. Such organic matter could originate as microbial and/or algal matter on the continents before the advent of vascular plants. This *in situ* production of CO_2 leads to the steep gradients in CO_2 concentration and $\delta^{13}\text{C}$ values in the upper 20–30 cm of the soil profile. Steady-state diffusive transport of CO_2 from the soil to the Earth's atmosphere constitutes a *de facto* two-end-member mixing system for soil CO_2 (ref. 3). The end members are CO_2 from oxidizing organic matter and CO_2 from the atmosphere^{3,15}. Application of this mixing model to the $\text{Fe}(\text{CO}_3)\text{OH}$ component in goethite as a proxy indicator of soil CO_2 leads to a mixing equation whose approximate mathematical linearity is accurate to about one part in ten thousand, because ${}^{13}\text{C} \ll {}^{12}\text{C}$:

$$\delta^{13}\text{C} = (\delta^{13}\text{C}_A - \delta^{13}\text{C}_O)[X(A)][1/X] + \delta^{13}\text{C}_O \quad (1)$$

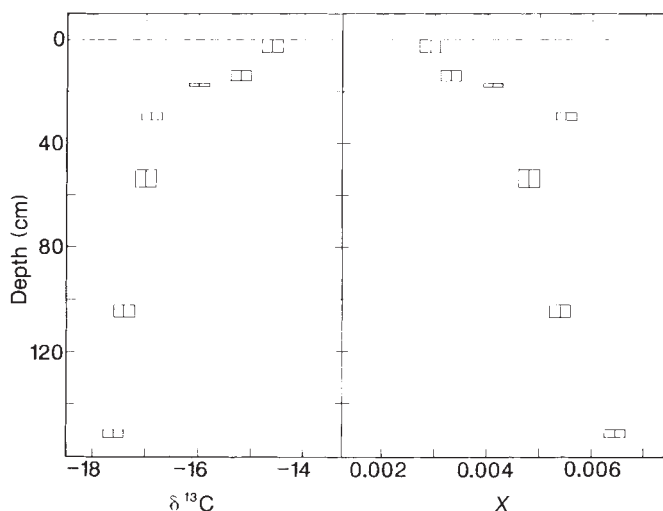


FIG. 1 Mole fraction (X) and $\delta^{13}\text{C}$ of the $\text{Fe}(\text{CO}_3)\text{OH}$ component in goethites of the Neda Formation as a function of depth below the Late Ordovician unconformity which bounds the upper surface of the deposit. Values of X and $\delta^{13}\text{C}$ of the $\text{Fe}(\text{CO}_3)\text{OH}$ component in a goethite sample are averages from measured plateau values for CO_2 incrementally evolved from goethite during stepwise, solid-state dehydration-decarbonation in vacuum at $\sim 230^\circ\text{C}$ (ref. 7). All samples are ground to $< 63\ \mu\text{m}$ and treated with concentrated H_2O_2 (30%) and 0.5–1.0 M HCl at room temperature before the stepwise vacuum dehydration. The incrementally evolved, plateau CO_2 does not originate from admixed organic matter or discrete carbonate mineral phases^{6,7}. Analytical precision of the mean for X is ± 0.0002 and for $\delta^{13}\text{C} \pm 0.2\text{‰}$ for CO_2 samples of $\sim 1\ \mu\text{mol}$. Samples from the upper $\sim 20\ \text{cm}$ are termed 'hard ore', and deeper samples are termed 'soft ore'⁸.

Ancient atmospheric CO_2 pressures inferred from natural goethites

Crayton J. Yapp & Harald Poths

Department of Geology, University of New Mexico, Albuquerque, New Mexico 87131, USA

THE role of changing atmospheric CO_2 concentrations in controlling global temperature can be investigated by examining variations in both CO_2 and climate preserved in the Earth's geological record. A model of the Earth's carbon cycle over the past 570 Myr suggests that, compared to its present value, the partial pressure of CO_2 (P_{CO_2}) may have been an order of magnitude higher in the early Palaeozoic, and about 4–6 times higher in the middle Mesozoic^{1,2}. Cerling³ used carbon isotope ratios in soil carbonate minerals to constrain atmospheric P_{CO_2} in portions of the Mesozoic and Cenozoic eras. Studies of the common mineral goethite ($\alpha\text{-FeOOH}$) have shown that it contains small quantities of a carbonate component ($\text{Fe}(\text{CO}_3)\text{OH}$), the concentration and carbon isotope content of which preserves a record of ambient P_{CO_2} at the time of formation^{4–7}. Here we present data for goethites from an ironstone in the Upper Ordovician Neda Formation (Wisconsin, USA)⁸, which suggest that 440 Myr ago atmospheric P_{CO_2} was ~ 16 times higher than today. However, this enhanced level of atmospheric CO_2 does not seem to have been accompanied by unusually warm temperatures in the tropics, and in fact may have been contemporaneous with high-latitude continental glaciation on Gondwanaland^{9,10}.

The upper surface of the Neda Formation ironstone is a Late Ordovician unconformity⁸. The measured mole fractions (X) and $\delta^{13}\text{C}$ values of the $\text{Fe}(\text{CO}_3)\text{OH}$ component in goethite samples collected over different depth intervals below the unconformity are listed in Table 1 and plotted as a function of sample depth in Fig. 1. There is a relatively abrupt increase in X and a decrease in $\delta^{13}\text{C}$ from the unconformity to a depth of $\sim 30\ \text{cm}$ (Fig. 1). At greater depths, the changes in X and $\delta^{13}\text{C}$ are more gradual. Carbon dioxide in modern soil profiles is characterized by a rapid concentration increase and $\delta^{13}\text{C}$ decrease from the surface to a depth of 20.00 cm with negligible changes in concentration and $\delta^{13}\text{C}$ at greater depths^{11–13}. If the variations in X and $\delta^{13}\text{C}$ of the $\text{Fe}(\text{CO}_3)\text{OH}$ component represent variations in the partial pressures and $\delta^{13}\text{C}$ values of CO_2 present at the time of goethite formation^{4,7}, the data in Fig. 1 suggest that the Neda Formation goethite was formed in a Late Ordovician weathering (soil) profile. A subaerial origin for the goethite in the Neda Formation is supported by the existence of the unconformity⁸ and by oxygen isotope data which indicate that meteoric water (originating as atmospheric precipitation) with a $\delta^{18}\text{O}$ value of $\sim -7\text{‰}$ was present at the time of goethite formation¹⁴. We know of no other hypothesis for the origin of the Neda

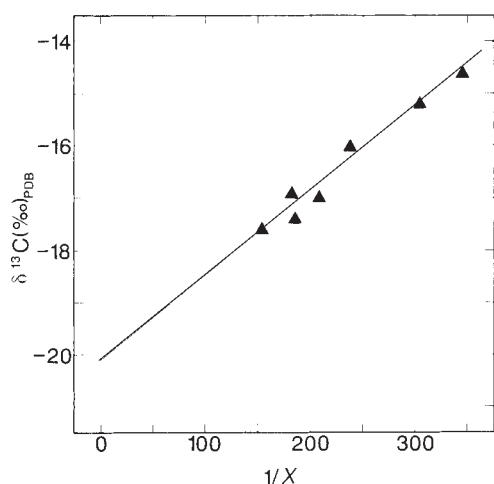


FIG. 2 $\delta^{13}\text{C}$ against $1/X$ for the $\text{Fe}(\text{CO}_3)\text{OH}$ component in the Neda Formation goethites of Table 1. The best-fit straight line through the data is depicted. The equation of the linear regression is $\delta^{13}\text{C} = 0.0162(1/X) - 20.1$, with $r = 0.98$. Linearity is expected if the goethite $\text{Fe}(\text{CO}_3)\text{OH}$ abundance and $\delta^{13}\text{C}$ values reflect ambient CO_2 in an ancient subaerial weathering profile.

$\delta^{13}\text{C}$ is the measured $\delta^{13}\text{C}$ value of the $\text{Fe}(\text{CO}_3)\text{OH}$ in goethite. $\delta^{13}\text{C}_A$ would be the $\delta^{13}\text{C}$ value for the $\text{Fe}(\text{CO}_3)\text{OH}$ component if atmospheric CO_2 were the only source of CO_2 in the weathering profile. $\delta^{13}\text{C}_O$ would be the $\delta^{13}\text{C}$ value in $\text{Fe}(\text{CO}_3)\text{OH}$ if CO_2 in the profile were due only to oxidation of organic matter. X is the measured mole fraction of $\text{Fe}(\text{CO}_3)\text{OH}$ in goethite. $X(A)$ would be the value of X if only atmospheric CO_2 were in the weathering profile. For constant $\delta^{13}\text{C}_A$, $\delta^{13}\text{C}_O$ and $X(A)$, the equation predicts a linear relationship between $\delta^{13}\text{C}$ and $1/X$. Figure 2 depicts a plot of $\delta^{13}\text{C}$ against $1/X$. Linear regression of the data of Fig. 2 yields

$$\delta^{13}\text{C} = 0.0162[1/X] - 20.1 \quad (2)$$

A high degree of linearity is reflected in the linear correlation coefficient (r) of 0.98. The slope of the line in equation (1) depends on the values of $\delta^{13}\text{C}_A$, $\delta^{13}\text{C}_O$ and $X(A)$. $\delta^{13}\text{C}_O$ is known from the intercept of equation (2). Therefore, a determination of the value of $X(A)$ from the slope of equation (2) requires independent knowledge of $\delta^{13}\text{C}_A$.

The sensitivity of $X(A)$ to the value of $\delta^{13}\text{C}_A$ was assessed by considering a 10% range for $\delta^{13}\text{C}_A$ with assigned extremum values of -6 and $+4\%$. From equations (1) and (2), the corresponding extremum values of $X(A)$ are 0.00115 and 0.00067. Yapp⁴ discussed a thermodynamically formulated relationship which shows how the calculated value of the atmospheric partial pressure P_{CO_2} would depend on knowledge of both $X(A)$ and the temperature (in degrees kelvin) of goethite formation. This equation is: $\log P_{\text{CO}_2} = \log X + 6.04 - 1,570/T$. The constant (6.04) is 0.21 smaller than the value reported by Yapp⁴, because of a new correction for nonstoichiometric water in the determination of X for sample NG-AtIRS-1 (see ref. 4). Ancient atmospheric P_{CO_2} values were calculated as a function of goethite formation temperature (T) using this relationship and plotted for the two extremes of $X(A)$ in Fig. 3. It is evident from Fig. 3 that independent knowledge of $\delta^{13}\text{C}_A$ (to determine $X(A)$) and T is essential if a reasonably constrained estimate of the Late Ordovician atmospheric P_{CO_2} is to be obtained.

Marine limestones and dolomites of the Late Ordovician have about the same $\delta^{13}\text{C}$ values as modern limestones¹⁶. If $\delta^{13}\text{C}$ values in atmospheric CO_2 vary in concert with marine carbonates³, then Late Ordovician $\delta^{13}\text{C}$ values in atmospheric CO_2 might have been $\sim -6.5\%$. Our results from other goethite

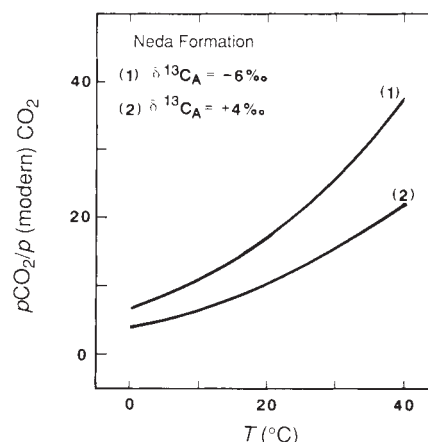


FIG. 3 Dependence of the calculated value of ancient atmospheric P_{CO_2} (as calculated from the $\delta^{13}\text{C}$ and X data for the Upper Ordovician Neda Formation) on the temperature of goethite formation, T ($^{\circ}\text{C}$). The $\delta^{13}\text{C}$ and X data used were those from the $\text{Fe}(\text{CO}_3)\text{OH}$ component in goethite ($\alpha\text{-FeOOH}$). Ancient atmospheric P_{CO_2} is normalized to modern P_{CO_2} . Curves 1 and 2 represent extremum $\text{Fe}(\text{CO}_3)\text{OH}$ $\delta^{13}\text{C}_A$ values, assuming a 10% range of end-member atmospheric CO_2 $\delta^{13}\text{C}$ (see text). At a goethite formation temperature of 24°C , and $\delta^{13}\text{C}_A$ of -1.5 , the Late Ordovician atmosphere might have had ~ 16 times more CO_2 than the modern atmosphere.

samples suggest that $\text{Fe}(\text{CO}_3)\text{OH}$ is enriched by $\sim 5\%$ in ^{13}C relative to gaseous CO_2 . Thus, the appropriate choice for $\delta^{13}\text{C}_A$ for the Neda Formation is $\sim -1.5\%$.

We recently obtained oxygen isotope data from goethite and phosphate in the Neda Formation ooids which yield a calculated temperature of goethite formation of $\sim 24^{\circ}\text{C} \pm 4^{\circ}\text{C}$. These data will be discussed elsewhere. Application of these $\delta^{13}\text{C}_A$ and T values to Fig. 3 yields a Late Ordovician atmospheric P_{CO_2} value which is ~ 16 times the modern value. Even if our results were in error by a factor of two, the goethite system would still indicate higher concentrations of atmospheric CO_2 in the Late Ordovician. This suggestion of higher atmospheric CO_2 partial pressures in the Early Palaeozoic is in general agreement with the model results of Berner^{1,2}.

The Neda Formation was at a latitude of $\sim 15^{\circ}\text{S}$ (ref. 17) and near sea level⁸ in the Late Ordovician. It is noteworthy that the apparently higher atmospheric P_{CO_2} in the Late Ordovician was associated with rather moderate low-altitude tropical temperatures as reflected in the goethite-phosphate oxygen isotope temperature of $\sim 24^{\circ}\text{C}$. Furthermore, the higher P_{CO_2} would have coincided with the existence of Late Ordovician continental ice sheets at high southern latitudes on Gondwanaland^{9,10}. At present, no palaeoclimatic models of the atmosphere-ocean system predict the development of continental ice sheets if atmospheric P_{CO_2} values were 16 times higher than modern¹⁸. Thus, if our interpretation of the Neda Formation data is correct, it suggests that the climatic models require modification, perhaps including consideration of variations in the value of the solar 'constant'.

Goethite-bearing oolitic ironstones (and laterites) of varying ages and locales are known from a substantial fraction of the Phanerozoic^{19,20}. Consequently, studies of the type discussed here could be undertaken to estimate atmospheric P_{CO_2} values in many important intervals in the Earth's past. $\square$

Received 23 August; accepted 5 November 1991.

1. Berner, R. A. *Science* **249**, 1382-1386 (1990).
2. Berner, R. A. *Am. J. Sci.* **291**, 339-376 (1991).
3. Cerling, T. E. *Am. J. Sci.* **291**, 377-400 (1991).
4. Yapp, C. J. *Chem. Geol.* **64**, 259-268 (1987).
5. Yapp, C. J. & Poths, H. *Clays Clay Miner.* **38**, 442-444 (1990).
6. Yapp, C. J. & Poths, H. *Geochim. cosmochim. Acta* **50**, 1213-1220 (1986).
7. Yapp, C. J. & Poths, H. *Geochim. Soc. Spec. Publ.* (in the press).

8. Paull, R. A. *Geology of Southeastern Wisconsin, 41st Tri-State Field Conf.* (ed. Nelson, K. G.) C-1-C-18 (1977).
9. Caputo, M. V. & Crowell, J. C., *Geol. Soc. Am. Bull.* **96**, 1020-1036 (1985).
10. Van Houten, F. B. *Paleogeogr. Paleoclimatol. Paleocool.* **80**, 245-254 (1990).
11. Cerling, T. E. *Earth planet. Sci. Lett.* **71**, 229-240 (1984).
12. Quade, J., Cerling, T. E. & Bowman, J. R. *Geol. Soc. Am. Bull.* **101**, 464-475 (1989).
13. Reardon, E. J., Allison, G. B. & Fritz, P. J. *Hydrol.* **43**, 355-371 (1979).
14. Yapp, C. J. *Geochim. cosmochim. Acta* **55**, 2627-2634 (1991).
15. Parada, C. B., Long, A. & Davis, S. N. *Isotope Geosci.* **1**, 219-236 (1983).
16. Holser, W. *Patterns of Change in Earth Evolution* (eds Holland, H. D. & Trendall, A. F.) 123-143 (Springer, Berlin, 1984).
17. Scotese, C. R., Bambach, R. K., Barton, C., Van Der Voo, R. & Ziegler, A. M. *J. Geol.* **87**, 217-277 (1979).
18. Crowley, T. J. & North, G. R. *Paleoclimatology* (Oxford University Press, New York, 1991).
19. Van Houten, F. B. & Bhattacharyya, D. P. *Ann. Rev. Earth planet. Sci.* **10**, 441-457 (1982).
20. Van Houten, F. B. *Climate in Earth History* 112-117 (National Academy, Washington DC, 1982).

ACKNOWLEDGEMENTS. This manuscript was typed by M. Sherman. We thank B. Witzke for providing IRMWIS-1, and T. Anderson for comments.

Spatial filtering precedes motion detection

M. J. Morgan

Laboratory for Neuroscience, Department of Pharmacology, University of Edinburgh Medical School, 1 George Square, Edinburgh, Scotland, UK

WHEN we perceive motion on a television or cinema screen, there must be some process that allows us to track moving objects over time: if not, the result would be a conflicting mass of motion signals in all directions. A possible mechanism, suggested by studies of motion displacement in spatially random patterns, is that low-level motion detectors have a limited spatial range, which ensures that they tend to be stimulated over time by the same object¹⁻³. This model predicts that the direction of displacement of random patterns cannot be detected reliably above a critical absolute displacement value ($D_{\max}$) that is independent of the size or density of elements in the display^{1,4,5}. It has been inferred that $D_{\max}$ is a measure of the size of motion detectors in the visual pathway^{1,3}. Other studies, however, have shown that D_{max} increases with element size⁶⁻⁸, in which case the most likely interpretation is that $D_{\max}$ depends on the probability of false matches between pattern elements following a displacement. These conflicting accounts are reconciled here by showing that $D_{\max}$ is indeed determined by the spacing between the elements in the pattern, but only after fine detail has been removed by a physiological prefiltering stage: the filter required to explain the data has a similar size to the receptive field of neurons in the primate magnocellular pathway. The model explains why $D_{\max}$ can be increased by removing high spatial frequencies from random patterns, and simplifies our view of early motion detection.

The upper motion displacement threshold ($D_{\max}$) was measured using random binary-luminance patterns (Fig. 1). The observers' task was to report the direction of displacement (up or down) on each trial. $D_{\max}$ was defined as the 80% correct point on the psychometric function⁹. The data (Fig. 2a) confirmed previous reports^{4,5} that $D_{\max}$ is independent of element size up to about 10 arcmin. Thereafter, however, $D_{\max}$ rose sharply, reaching levels of more than 1 degree. Similar findings have been reported by Sato¹⁰. The pattern of results (a flat region followed by a linear increase) would be expected if the random patterns were spatially filtered before the displacement of edges (or other features) were detected, and if the flat region corresponded to a range of element sizes that were smaller than or similar in size to the filter⁷. The effect of the filter size is illustrated in Fig. 3, which shows: a random binary pattern (a); a Mexican-hat-shaped filter similar to the cross-sectional receptive field profile of retinal ganglion cells, lateral geniculate cells, and simple cortical cells^{11,12} (b); and the convolution between the pattern and the filter (c). This is shown for three

different element sizes. The filtered profile is very similar for the two smaller element sizes but markedly different for the largest element size.

To measure the effect of different sizes of filter and pattern element, I computed the mean distance between like-signed zero-crossings in profiles such as those in Fig. 3c. The behaviour of the model is illustrated in Fig. 2b: it captures both the flat region and the region of linear increase found in the psychophysical data. The assumption in the model is that the greater

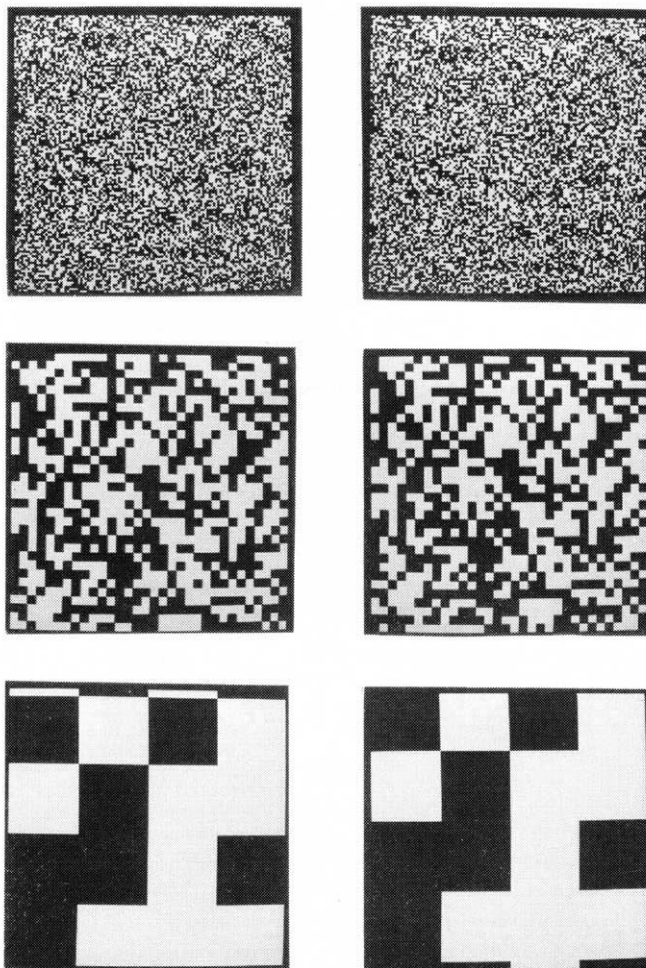


FIG. 1 Each row shows a pair of patterns that produces a perception of motion if they are presented successively in the same location. Three different element sizes are illustrated. In the experiment, the observer's task was to report the direction of motion (up or down), which was randomly varied in magnitude and direction over trials to determine $D_{\max}$, the upper displacement limit for motion detection. To determine the threshold for each element size, eight sizes of displacement were randomly interleaved until the observer had seen ten trials of each displacement; from these data a psychometric function was constructed and the 80% correct level determined⁹. Each threshold was determined four times for a range of element sizes 1.2-72.0 arcmin. The observers were the author (M.M.), another experienced observer (M.F.) and two others with little previous experience of psychophysical tasks (J.H. and L.P.). In the experiment, element size was varied over a range 1-72 arcmin and $D_{\max}$ was determined for each size. The result was that $D_{\max}$ was independent of element size for sizes less than about 10 arcmin, but strongly dependent thereafter (Fig. 2). A simple filtering model (Figs 2 and 3) accounts for the data. Patterns were generated on a ManiTron monochrome display at an 85 Hz frame rate, and each of the two patterns was presented for 70 ms, with a single blank frame (12 ms) interval between frames. The patterns subtended a visual angle of 5×5 deg at the standard viewing distance by 2.28 m. The contrast of the pattern was 100% in the dark but this was reduced to 50% by ambient room lighting, which supplied a veiling luminance on the screen of about 50 cd m^{-2} .

the distance between like-signed zero-crossings, the larger the displacement that can be detected without ambiguity by matching nearest neighbour zero-crossings between the two patterns (a similar idea has been applied to stereoscopic matching¹³). Zero-crossings were chosen for convenience, without any implication that they are the actual primitives used in motion matching: other possibilities such as peaks, and the distance between the centroids of zero-bounded masses¹⁴ gave very similar results. The separation between zero-crossings is computed only in the direction orthogonal to the motion: this simplification was justified by a subsidiary experiment with rectangular elements, which showed that elongating the elements at right angles to the direction of motion had no effect upon $D_{\max}$. For large element sizes the mean distance between like zero-crossings is four times the size of the pattern elements, as would be predicted from the statistics of a binary pattern independently of the filter size, and the curves intercept the y -axis when the mean distance between like-signed zero-crossings is equal to the filter space

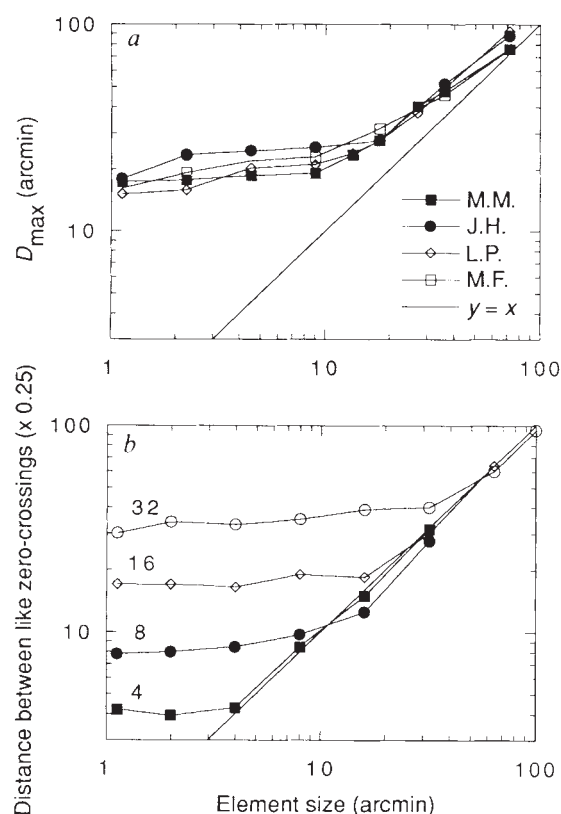


FIG. 2 Results of an experiment and a model to determine the upper limit for $D_{\max}$ as a function of the size of elements in random patterns illustrated in Fig. 1. *a*, Results of the experiment where each set of points represents the data from a different observer. Each point is the mean of four readings. $D_{\max}$ is nearly constant over a range 1–10 arcmin, but arises thereafter. *b*, Model to account for the shape of the curves in (*a*). Each point shows the mean of 10 independent computations based on different random stimulus profiles. Each curve shows the results for a different laplacian-of-a-gaussian filter, specified by the standard deviation of the gaussian. The diagonal straight line shows the relation $y=x$, where y is one quarter of the mean distance between like-signed zero-crossings. The justification for this measure is that beyond 1/4 of the distance between zero crossings, motion displacements become increasingly ambiguous. The psychophysical data also appear to converge on a unit slope between $D_{\max}$ and element size. Note that the model curves capture the flat portion of the psychophysical curves at small element sizes. Taking into account the knee-point of the curves and their y intercept, a physiological filter with a standard deviation in the region of 8–16 arcmin would be inferred.

constant, the ideal result for filtered gaussian noise (N. Barton, personal communication).

The knee-point on the curves occurs when the standard deviation of the filter is roughly the same as the element size. Using this fact, we can estimate the standard deviation of the physiological filter to be in the region 8–16 arcmin. For the case of a laplacian-of-gaussian filter this corresponds to a receptive field centre size of 16–32 arcmin. This is probably too large to characterize foveal cells of the parvocellular pathway, but it is not incompatible with estimates of magnocellular (M) units, believed on other grounds to form the substrate of primate motion perception¹⁵. One study, for example, described few M cells in macaque monkeys as having centres smaller than 10 arcmin, with many greater than 20 arcmin¹⁶.

The convolutions illustrated in Fig. 3 were normalized in height: their amplitude diminished when the element size was smaller than the filter. This may explain the small but systematic trend in the psychophysical data towards a reduction in performance at the smallest element sizes. That this decline is not greater may well reflect the comparatively high contrast sensitivity of magnocellular neurons^{15,16}.

The model does not claim that motion detection is limited to displacements smaller than the filter: indeed the data show orderly detection of displacements five times greater than the centre diameter. Rather, the pattern is prefiltered before the actual detection of motion occurs at a higher level, and the only limit to the matching process is in the pattern itself. The model shares with the 'elaborated Reichardt model' a localized spatial filtering preceding motion detection¹⁷, but it does not involve the additional assumption of that model that there is a fixed range of bilocal detector separations, matched to the size of the prefilter. The traditional $D_{\max}$ value of about 15 arcmin for foveal vision reflects, not the range over which correlations can be detected, but the small sizes of element that were originally used to determine the motion threshold¹, in combination with a relatively coarse prefilter.

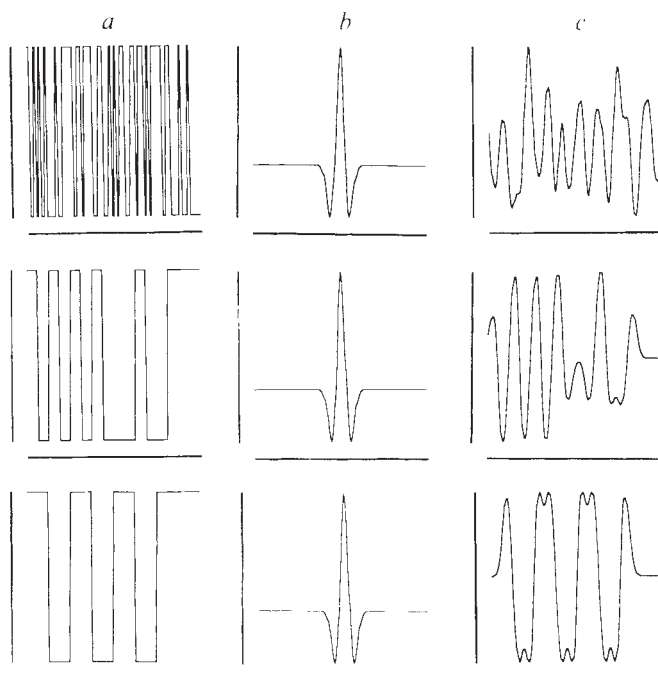


FIG. 3 *a*, Result of convolving the stimulus patterns in *a* with the filter profile in *b*. The pattern represents the luminance cross-section of a random binary pattern like those in Fig. 1. Three different element sizes are illustrated (top, middle and bottom rows). Note that the spacing between zero-crossings does not differ between the top and middle stimuli, but the spacing in the bottom row is larger.

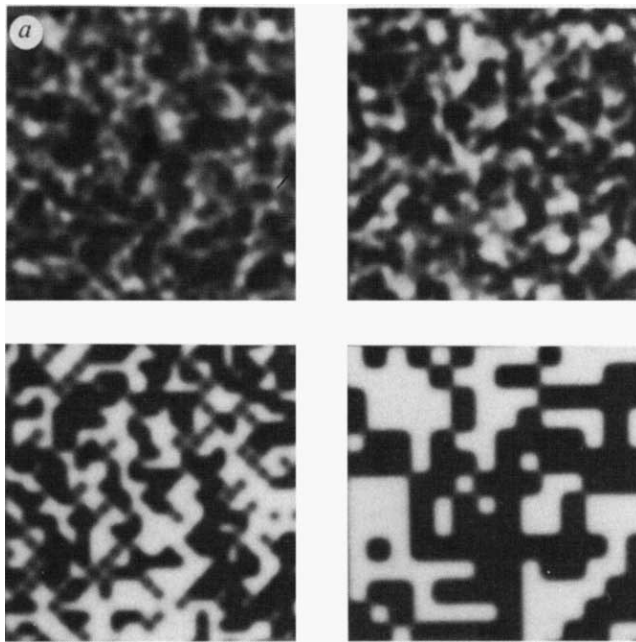
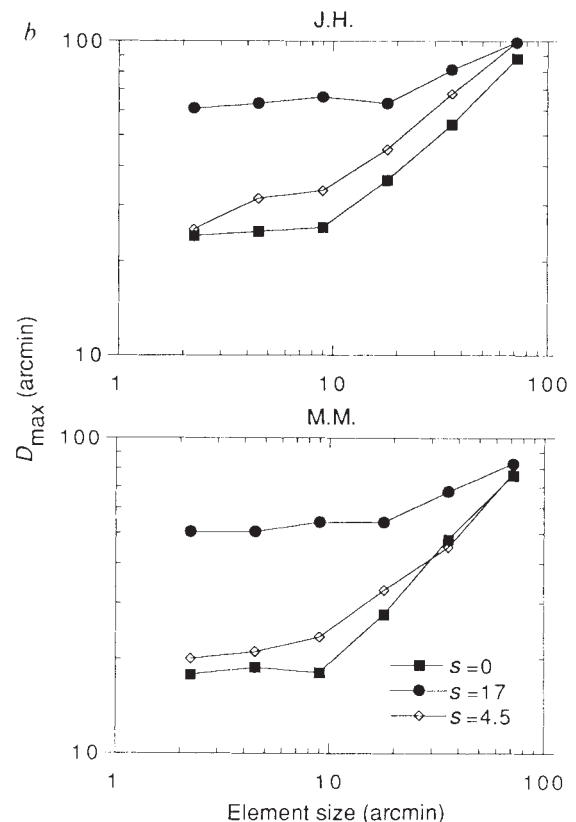


FIG. 4 *a*, Examples of low-pass filtered random binary patterns used in an experiment to measure $D_{\max}$ as a function of original element size and filter size. The patterns were generated by convolving binary patterns with pixel size 1 (top left), 4 (top right), 8 (bottom left) and 16 (bottom right) with a Gaussian filter having a standard deviation of 3.8 pixels. The original pattern had dimensions 256×256 pixels. Note that the blob size and separation after filtering is little affected by the original pixel size, except as the largest size. This leads to the prediction that the upper motion displacement threshold $D_{\max}$ will be little affected by element size, until the element size is large with respect to the filter: this prediction was confirmed by the



experiment result in *b*. *b*, Results of an experiment in which $D_{\max}$ was measured as a function of element size with different amounts of prefiltering of the pattern. The data from a standard deviation of zero ($s=0$) (unfiltered) are the original data from Fig. 2. For $s=4.5$ and $s=17$ the filters were Gaussian with standard deviations of 4.5 and 17 arcmin, respectively. Results are shown for two observers (J.H. and M.M.): note that for both observers, the smaller filter has little effect on the shape of the curve, whereas the larger filter extends the region over which $D_{\max}$ is invariant with original element size. The results may be compared with the theoretical curves in Fig. 2*a*.

The model predicts that low-pass spatial frequency filtering (Fig. 4*a*) will extend the region over which $D_{\max}$ is invariant with element size. If we apply a filter before the putative physiological filter, the effect should be to shift the $D_{\max}$ /element size function to a new curve in the family in Fig. 2*b*. The experimental results in Fig. 4*b* show that this prediction is correct. The patterns were filtered by convolution with a Gaussian mask, and then normalized to have the same peak-to-trough contrast as the original unfiltered patterns. A filter with a standard deviation (s.d.) of 4.5 min, which is smaller than the putative physiological filter, has little effect on the knee-point on the $D_{\max}$ /element size curve, but a larger filter (s.d. = 17') clearly extends the flat region as well as increasing $D_{\max}$ overall.

This analysis solves a theoretical puzzle about motion percep-

tion. The displacement limit in band-limited patterns is roughly proportional to their centre spatial frequency, as would be expected from the zero-crossing model used here^{18,19}. This has been explained by supposing that there are multiple motion filters at different spatial scales, each with its own $D_{\max}$ value¹⁹. But the problem then is to explain why $D_{\max}$ values for broad-band patterns, such as those used to determine $D_{\max}$ traditionally, do not reflect the undoubted presence of low spatial frequency components in those patterns. To solve this difficulty it has been proposed¹⁹ that high spatial frequency filters inhibit low spatial frequencies. The need for this elaborate scheme disappears, however, if the detection limit for both broad- and narrow-band patterns is set by the mean spatial interval between their elements, albeit after some preliminary spatial filtering. □

Received 5 August; accepted 6 November 1991.

1. Braddick, O. J. *Vision Res.* **14**, 519–527 (1974).
2. Anstis, S. M. *Phil. Trans. R. Soc. B* **290**, 153–168 (1980).
3. Braddick, O. J. *Phil. Trans. R. Soc. B* **290**, 137–151 (1980).
4. Baker, C. L. & Braddick, O. J. *Vision Res.* **22**, 1253–1260 (1982).
5. Sato, T. *Vision Res.* **29**, 1749–1758 (1989).
6. Lappin, J. S. & Bell, H. H. *Vision Res.* **16**, 161–168 (1976).
7. Cavanagh, P., Boeglin, J. & Favreau, O. *Perception* **14**, 151–162 (1985).
8. Chang, J. J. & Julesz, B. *Vision Res.* **23**, 639–646 (1983).
9. Morgan, M. J. & Cleary, R. F. *Vision Res.* (in the press).
10. Sato, T. *Perception* **19**, A7 (1990).
11. Rodieck, R. W. & Stone, J. J. *Neurophysiol.* **28**, 883–849 (1965).

12. Marr, D. & Hildreth, E. C. *Proc. R. Soc. B* **207**, 187–217 (1980).
13. Marr, D. & Poggio, T. *Proc. R. Soc. B* **204**, 301–328 (1979).
14. Watt, R. J. & Morgan, M. J. *Vision Res.* **25**, 1661–1674 (1985).
15. Schiller, P. H., Logothetis, N. K. & Charles, E. R. *Nature* **343**, 68–69 (1990).
16. Crook, J. M., Lange-Malecki, B., Lee, B. B. & Valberg, A. *J. Physiol. Lond.* **396**, 205–224 (1988).
17. Van Santen, J. P. H. & Sperling, G. *J. opt. Soc. Am.* **1**, 451–473 (1984).
18. Bischof, W. F. & DiLollo, V. *Vision Res.* **30**, 1341–1362 (1990).
19. Cleary, R. F. & Braddick, O. J. *Vision Res.* **30**, 317–327 (1990).

ACKNOWLEDGEMENTS. This work was supported by the SERC. I thank N. Barton for help in deriving the statistics of zero-crossings in random patterns and L. Partridge, G. Mather and O. Braddick for comments.

Strong selection on a gene that influences reproductive competition in a social insect

Kenneth G. Ross

Department of Entomology, University of Georgia, Athens, Georgia 30602, USA

COMPETITION among nestmate females for reproductive opportunities is a fundamental property of insect societies, the outcome of which defines the form of colony social organization and the nature of subsequent social evolution¹⁻³. Several factors influencing the outcome of competition among potential reproductives have been identified in eusocial Hymenoptera, including age⁴, size⁵⁻⁷ and degree of ovary or exocrine gland development⁷⁻¹⁰. In no case is the genetic basis of traits associated with success in these social contests understood, although there is a heritable component to their expression in laying worker honey bees¹¹. I report here the existence of a single mendelian factor that strongly influences success in reproductive competition in multiple-queen societies of

an ant, and I also propose a mechanism by which variation is maintained at this gene despite the presence of strong directional selection.

The fire ant *Solenopsis invicta* is an advanced eusocial insect in which two distinct forms of social organization exist. Colonies in monogyne populations are simple families headed by a single reproductive queen, whereas colonies in polygyne populations contain up to several hundred reproductive queens^{12,13}. Initiation of reproduction by queens occurs through different routes in the two social forms. Monogyne queens found new nests independently, either alone or in groups, but never with workers in attendance¹⁴. By contrast, polygyne queens initiate reproduction in their natal or other polygyne nest in the company of workers and older queens^{15,16}.

An electrophoretic survey of wingless reproductive queens in a polygyne population from northern Georgia, United States, revealed consistent large-scale departures of genotype proportions from those expected under Hardy-Weinberg equilibrium (HWE) at the enzyme locus *Pgm-3*. Most striking was the complete absence of one genotypic class among these queens over both years of the survey (Table 1). These results are not artefacts of interpretation because extensive family studies show that this marker is inherited as the product of a single mendelian locus

TABLE 1 Observed and expected population genotype and allele frequencies and fit to Hardy-Weinberg genotype proportions at the locus *Pgm-3* in *S. invicta*

	Genotype frequencies observed (expected)			Allele frequencies		Deviations from Hardy-Weinberg genotype proportions
	<i>aa</i>	<i>ab</i>	<i>bb</i>	<i>a</i>	<i>b</i>	
Polygyne population						
Queens						
Wingless reproductives						
1990						
mated (<i>N</i> =28, <i>n</i> =760)	0 (0.139)	0.746 (0.468)	0.254 (0.393)	0.373	0.627	0.98
virgin (<i>N</i> =21, <i>n</i> =113)	0 (0.170)	0.827 (0.485)	0.173 (0.345)	0.413	0.587	0.98
all (<i>N</i> =28, <i>n</i> =873)	0 (0.143)	0.756 (0.470)	0.244 (0.387)	0.378	0.622	0.94
1991						
mated (<i>N</i> =38, <i>n</i> =307)	0 (0.138)	0.744 (0.468)	0.256 (0.394)	0.372	0.628	1.0
virgin (<i>N</i> =24, <i>n</i> =57)	0 (0.112)	0.670 (0.446)	0.330 (0.442)	0.335	0.665	0.83
all (<i>N</i> =52, <i>n</i> =467)	0 (0.138)	0.744 (0.468)	0.256 (0.394)	0.372	0.628	1.0
Winged nonreproductives						
1990 (<i>N</i> =28, <i>n</i> =886)	0.212 (0.288)	0.649 (0.497)	0.139 (0.215)	0.537	0.463	0.20
1991 (<i>N</i> =27, <i>n</i> =446)	0.099 (0.193)	0.680 (0.493)	0.221 (0.314)	0.439	0.561	0.30
Workers						
Pupae (<i>N</i> =28, <i>n</i> =922)	0.194 (0.272)	0.655 (0.499)	0.151 (0.229)	0.521	0.479	0.24
Adults (<i>N</i> =28, <i>n</i> =724)	0.263 (0.320)	0.606 (0.491)	0.131 (0.189)	0.566	0.434	0.20
Males (<i>N</i> =22, <i>n</i> =112)	—	—	—	0.251	0.749	—
Monogyne population						
Queens						
Wingless reproductives						
mature mother queens (<i>n</i> =59)	0.542 (0.582)	0.441 (0.362)	0.017 (0.056)	0.763	0.237	n.s. (<i>P</i> >0.10)
newly-mated queens (<i>n</i> =271)	0.542 (0.555)	0.406 (0.380)	0.052 (0.065)	0.745	0.255	n.s. (<i>P</i> >0.10)
Winged nonreproductives (<i>N</i> =65, <i>n</i> =601)	0.476 (0.498)	0.459 (0.416)	0.065 (0.086)	0.706	0.294	0.04
Workers						
Pupae (<i>N</i> =52, <i>n</i> =1316)	0.575 (0.573)	0.362 (0.368)	0.063 (0.059)	0.757	0.243	0
Adults (<i>N</i> =39, <i>n</i> =160)	0.489 (0.466)	0.388 (0.433)	0.123 (0.101)	0.683	0.317	0
Males (<i>n</i> =108)	—	—	—	0.815	0.185	—

Pgm-3 encodes the enzyme phosphoglucosyltransferase (EC 5.4.2.2). *N*, Number of nests; *n*, number of individuals studied. In cases where only *n* is shown, only one individual was sampled per nest (monogyne males, monogyne mother queens) or individuals were not collected in association with nests and are assumed to each originate from a different nest (newly mated monogyne queens). All individuals of a single class were collected during a single season and, unless otherwise indicated, a single year. Observed genotype and allele frequencies represent the means of 50 random draws (with replacement) of single genotypes from each nest in cases where more than one individual was sampled per nest. This procedure ensures the independence of sampled genotypes in strongly family-structured social insect populations (see for example ref. 26). Genotype proportions expected under Hardy-Weinberg equilibrium are shown in parentheses. Deviations from Hardy-Weinberg genotype proportions are either the proportion of the 50 random draws in which the drawn genotypes departed significantly from ratios expected under Hardy-Weinberg equilibrium (cases where >1 individual sampled per nest) or are the results of a single test for departure of the entire sample from Hardy-Weinberg equilibrium (monogyne wingless reproductives) (χ^2 test with Yates' continuity correction²⁷, 5% significance level). Haplotype polygyne males were distinguished from their more common diploid (infertile) male nestmates on the basis of banding patterns at four polymorphic enzyme loci (*Est-4*, *G3pdh-1*, *Pgm-1*, *Pgm-3*) and by their smaller size²⁴. Haplotype male genotype frequencies are the same as allele frequencies. Genotypes at *Pgm-3* were scored using the same electrophoretic methods that were used for *Pgm-1* (see Table 3).

TABLE 2 Observed and expected genotype classes at the locus *Pgm-3* in simple families of *S. invicta*

	<i>aa</i>	Genotype classes observed (expected)			<i>bb + ab</i>	Deviations from 50:50 ratio for split genotype classes
		<i>ab</i>	<i>bb</i>	<i>aa + ab</i>		
Monogyne population						
winged nonreproductive queens ($N=33$, $n=574$)	11 (11.6)	7 (6.9)	0 (0.7)	9 (9.7)	6 (4.1)	0
worker pupae ($N=52$, $n=1316$)	23 (22.6)	7 (9.6)	0 (0.7)	16 (14.5)	6 (4.6)	0
Polygyne population						
worker pupae ($N=69$, $n=766$)	0 (0)	14 (12.1)	2 (5.1)	39 (36.4)	14 (15.4)	0.06

N , Number of families; n , number of individuals studied (from eight to 54 offspring were genotyped for each family). Data for the monogyne population come from colonies collected in the field (such colonies are simple families¹²). Data for the polygyne population were obtained by collecting 2–4 wingless reproductive queens from each of 25 randomly selected polygyne field nests and isolating these queens in laboratory rearing units for at least 6 weeks, at which time pupae collected from the units were known to be the resident queens' offspring. The genotypes of the mother queens were obtained in addition to offspring genotypes for all 69 of these polygyne families, as well as for 31 of the 85 monogyne families. Values shown are the observed numbers of families with particular classes of genotypes; the expected numbers of such families (in parentheses) were calculated from the population genotype frequencies for the two sexes and consequent expected frequencies of *Pgm-3* mating types. All matings in the monogyne population and 80% of matings in the polygyne population were assumed to be by monogyne males (see text and Fig. 1). Evidence for mendelian inheritance of the product of *Pgm-3*: Only one or two genotype classes were observed among female offspring in single monogyne or polygyne families. When two genotype classes were present, one was always the heterozygote class and the two classes typically segregated in a 50:50 ratio (the proportion of significant deviations from this ratio, estimated from the binomial distribution at the 5% probability level, is given in the final column). The mother's genotype, when available, was always consistent with the offspring genotypes being the result of a single mating¹². The close correspondence between the observed and expected numbers of families with particular genotype classes, the identities of classes within families, the segregation ratios, and the concordance of mother/offspring genotypes demonstrate that *Pgm-3* behaves as a single mendelian locus.

(Table 2). The genotypic disequilibrium observed at *Pgm-3* in polygyne queens disagrees with results from all other polymorphic enzyme loci surveyed in the same polygyne population; genotype distributions for these 10 additional markers show close correspondence to HWE (Table 3). Disequilibrium at *Pgm-3* therefore arises from forces acting uniquely on this or a closely linked gene rather than from processes related to the breeding or dispersal systems, which are expected to affect all genomic markers similarly.

Strong selection on *Pgm-3* in the context of reproductive competition provides the most plausible explanation for the observed genotypic disequilibrium. Although all classes of females in the polygyne population show disequilibrium at *Pgm-3*, specific genotype proportions and the consequent levels of disequilibrium differ considerably between reproductive and nonreproductive females (Table 1). Reproductive queens lack the homozygous *Pgm-3^a/3^a* genotype, and the unusual ratios of the other two genotypes create extreme genotypic disequilibrium in these reproductives. By contrast, all three genotypes are found among nonreproductive polygyne females (winged queens and workers), although consistent excesses of heterozygotes create moderate levels of disequilibrium. It is important, that genotype and allele frequencies in the monogyne form are similar among all classes of females and differ strongly from proportions in the polygyne form, and monogyne genotype proportions are in HWE.

The absence of *Pgm-3^a/3^a* homozygotes among reproductive polygyne queens and their presence among nonreproductive queens suggest that selection acts by eliminating queens with this genotype before they are recruited into a colony's pool of reproductives. This view is consistent with the similar ratios of *ab* to *bb* genotypes in reproductive queens (3.0:1) and the winged nonreproductive queens from which they originate (3.9:1). Also, preliminary data indicate a greatly diminished survival probability for *Pgm-3^a/3^a* queens relative to the other genotypes when these queens initiate reproduction in isolation and are reintroduced into their natal polygyne nests (L. Keller and K.G.R., manuscript in preparation). That selection acting on polygyne queens does not result from wholesale breakdown of adult viability is indicated by the fact that *Pgm-3^a/3^a* is the most common genotype among reproductive queens in mature monogyne colonies (Table 1).

Selection on *Pgm-3* seems to be related to some unique feature of the process by which polygyne queens of *S. invicta* become functionally reproductive. Genotype proportions in monogyne queens remain relatively unchanged from the time before their

mating flights, to just after mating, to when they are reproductives in large colonies (Table 1), indicating no comparable type of selection in this social form. The most conspicuous difference between the two forms in the process of reproductive development is that it is initiated in the presence of fully functional reproductive queens in the polygyne but not the monogyne form, a difference likely to have important consequences for reproductive competition. Reproductive competition among polygyne fire ant queens is mediated largely by workers, who seem actively to select the queens that become reproductives while destroying the remainder, most probably in response to each queen's pheromonal output^{8,17–19}. This study demonstrates that a queen's genotype at a single mendelian locus is a determinant of the outcome of this competitive process, suggesting that the gene product may be involved in regulating production of a queen control pheromone.

Complete loss of reproductive opportunities for polygyne

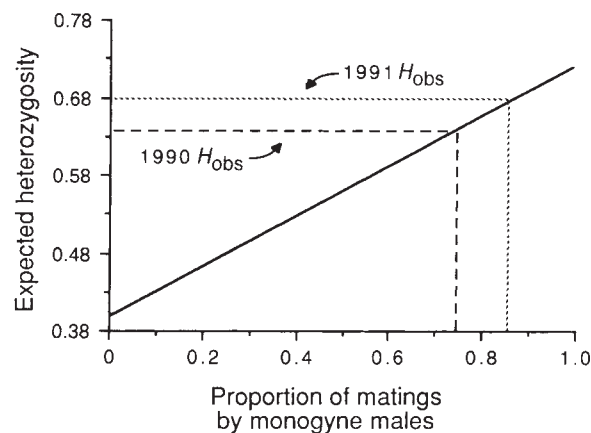


FIG. 1 Relationship between proportion of matings of polygyne queens attributed to monogyne males and expected heterozygosity of female offspring at *Pgm-3*. The solid line was generated from equation 2.15 of ref. 20, which estimates values of heterozygosity expected under male haploidy when allele frequencies differ between the sexes. Values are based on the average allele frequencies for polygyne reproductive queens over both years of the study and on the allele frequencies expected for their male mates when monogyne males are responsible for varying proportions of successful matings (haploid polygyne males being responsible for the remainder). The average heterozygosities observed for nonreproductive females for each year of the study (H_{obs}) correspond to heterozygosities expected when 74 or 85% of matings by polygyne queens are attributed to monogyne males.

TABLE 3 Deviations from Hardy-Weinberg genotype proportions at 10 polymorphic loci in polygynous *S. invicta*

	Aat-2	Acoh-1	Acoh-5	Acy1	Ddh-1	Est-4	Est-6	G3pdh-1	Pgm-1	Pro-5
Workers										
larvae (N = 31, n = 1012)	0	—	—	—	0.02	0	—	—	0	—
pupae (N = 31, n = 1015)	—	0	0	—	—	—	0	—	0.04	—
adults (N = 31, n = 976)	0	—	—	—	—	0.04	—	0	0	—
Queens										
wingless reproductives (N = 31, n = 616)	0	0	0	—	—	0.02	—	0	0.02	—
winged nonreproductives (N = 31, n = 954)	0.02	0	—	0	—	0.04	—	0.02	0.04	0.02

N, Number of nests; n, mean number of individuals studied per locus for each class of female. Values represent the proportion of 50 random draws of single genotypes per nest in which the ratios of the drawn genotypes departed significantly from those expected under Hardy-Weinberg equilibrium (χ^2 test with Yates' continuity correction, 5% significance level; see Table 1 legend). Dashes indicate that the markers were not scored in a particular material. Genotypes were scored following horizontal electrophoresis in 14% starch gels using a pH 6.0 amine-citrate (morpholine) continuous buffer system (*Ddh-1*, *Est-4*, *Est-6*, *G3pdh-1*, *Pgm-1*, *Pro-5*), a pH 8.6 Tris-borate-EDTA continuous buffer system (*Aat-2*, *Acy1*), or a pH 6.5 Tris-citrate continuous buffer system (*Acoh-1*, *Acoh-5*), with the protein bands visualized by specific histochemical staining^{12,28,29}. Mendelian inheritance of all 10 markers has been confirmed by family studies (ref. 12; D. Shoemaker, J. Costa and K.G.R., manuscript in preparation).

queens homozygous for *Pgm-3^a* suggests that the dynamics of this allele should mimic those of a recessive lethal, resulting in virtual monomorphism for the alternate allele in relatively few generations in the absence of other evolutionary forces²⁰. Yet *Pgm-3^a* was common in the polygynous population over both years of this study. Because a high frequency of this allele in polygynous males can be ruled out as an explanation for the maintenance of this polymorphism (Table 1), the most likely cause is that *Pgm-3^a* is continually reintroduced into the polygynous population through dispersing monogynous males. *Pgm-3^a* occurs at high frequency in these males (Table 1), and their ability to disperse widely²¹ and the modest size of the polygynous population studied²² allow ample opportunity for males from surrounding monogynous populations to reach polygynous females. Moreover, the *Pgm-3^a* allele is sufficiently more common in monogynous males than in polygynous reproductive queens that substantial excess heterozygosity should be generated at *Pgm-3* if they interbreed²⁰, as is observed. Indeed, the excess heterozygosity found in nonreproductive polygynous females can be fully accounted for if some 80% of matings in the polygynous population involve monogynous males (Fig. 1). Additionally, the distribution of genotype classes observed in 69 families derived from single polygynous queens closely matches the distribution expected if monogynous males are responsible for 80% of the matings (Table 2). Finally, because of reduced allelic variation at the putative sex-determining locus, 80–95% of males produced in polygynous populations of *S. invicta* are infertile diploids, whereas males in monogynous populations are invariably haploid^{23–25}. Thus there may be too few fertile haploid males produced in polygynous populations to compete effectively with immigrant monogynous males for fertilizations.

This study provides evidence that a single gene can strongly influence the outcome of reproductive competition in a social insect, that this effect is dependent on social context, and that both selection and gene flow are important in determining genotype and allele frequencies at this gene in a specific social environment. Identification of the gene and physiological role of its product may elucidate the manner in which its phenotypic social effects are achieved and so provide a model for the genetic basis of queen competition and control in insect societies. □

11. Moritz, R. F. A. & Hillesheim, E. *Behav. Ecol. Sociobiol.* **17**, 87–89 (1985).
12. Ross, K. G. & Fletcher, D. J. C. *Behav. Ecol. Sociobiol.* **17**, 349–356 (1985).
13. Vargo, E. L. & Fletcher, D. J. C. *Physiol. Entomol.* **12**, 109–116 (1987).
14. Markin, G. P., Collins, H. L. & Dillier, J. H. *Ann. ent. Soc. Am.* **65**, 1053–1058 (1972).
15. Glancey, B. M. & Lofgren, C. S. *Fla. Ent.* **71**, 581–587 (1988).
16. Vargo, E. L. & Porter, S. D. *Ann. ent. Soc. Am.* **82**, 307–313 (1989).
17. Fletcher, D. J. C. & Blum, M. S. *Science* **212**, 73–75 (1981).
18. Fletcher, D. J. C. in *Advances in Invertebrate Reproduction 4* (eds Porchet, M., Andries, J.-C. & Dhainaut, A.) 305–316 (Elsevier, Amsterdam, 1986).
19. Vargo, E. L. & Ross, K. G. *J. Insect Physiol.* **35**, 587–593 (1989).
20. Hedrick, P. W. *Genetics of Populations* (Jones & Bartlett, Boston, 1985).
21. Markin, G. P., Dillier, J. H., Hill, S. O., Blum, M. S. & Hermann, H. R. *J. Georgia ent. Soc.* **6**, 145–156 (1971).
22. Fletcher, D. J. C. *J. Georgia ent. Soc.* **18**, 538–543 (1983).
23. Hung, A. C. F., Vinson, S. B. & Summerlin, J. W. *Ann. ent. Soc. Am.* **67**, 909–912 (1974).
24. Ross, K. G. & Fletcher, D. J. C. *Evolution* **39**, 888–903 (1985).
25. Ross, K. G. & Fletcher, D. J. C. *Behav. Ecol. Sociobiol.* **19**, 283–291 (1986).
26. Crozier, R. H., Smith, B. H. & Crozier, Y. C. *Evolution* **41**, 902–910 (1987).
27. Weir, B. S. *Genetic Data Analysis* (Sinauer, Sunderland, Massachusetts, 1990).
28. Qavi, H. & Kit, S. *Biochem. Genet.* **18**, 669–679 (1980).
29. Murphy, R. W., Sites, J. W., Buth, D. G. & Haufier, C. H. in *Molecular Systematics* (eds Hillis, D. M. & Moritz, C.) 45–126 (Sinauer, Sunderland, Massachusetts, 1990).

ACKNOWLEDGEMENTS. I thank L. Keller, M. A. Moran, D. Shoemaker, J. Cook, M. Arnold, and J. Costa for advice and assistance. This work was supported in part by the National Geographic Society and the Georgia Agricultural Experiment Stations, University of Georgia.

Stationary and drifting spiral waves of excitation in isolated cardiac muscle

Jorge M. Davidenko, Arcady V. Pertsov, Remy Salomonsz, William Baxter & José Jalife

Department of Pharmacology, SUNY Health Science Center, 750 East Adams Street, Syracuse, New York 13210, USA

EXCITABLE media can support spiral waves rotating around an organizing centre^{1–7}. Spiral waves have been discovered in different types of autocatalytic chemical reactions^{8,9} and in biological systems^{10–12}. The so-called 're-entrant excitation' of myocardial cells¹³, causing the most dangerous cardiac arrhythmias, including ventricular tachycardia and fibrillation, could be the result of spiral waves^{1,2}. Here we use a potentiometric dye^{14,15} in combination with CCD (charge-coupled device) imaging technology^{16,17} to demonstrate spiral waves in the heart muscle. The spirals were elongated and the rotation period, T_s , was about 180 ms (3–5 times faster than normal heart rate). In most episodes, the spiral was anchored to small arteries or bands of connective tissue, and gave rise to stationary rotations. In some cases, the core drifted away from its site of origin and dissipated at a tissue border. Drift was associated with a Doppler shift in the local excitation period, T , with T ahead of the core being about 20% shorter than T behind the core.

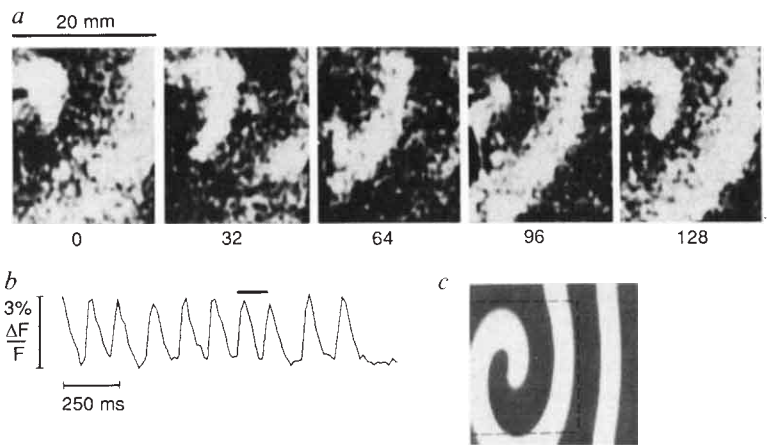
Sustained spiral-wave activity was consistently initiated in thin (20 × 20 × 0.5 mm) slices of sheep and dog epicardial muscle by crossfield stimulation^{2,15,18}. The rotation period ($T_s = 183 \pm$

Received 8 August; accepted 4 November 1991.

1. Wilson, E. O. *The Insect Societies* (Belknap, Cambridge, Massachusetts, 1971).
2. West-Eberhard, M. J. in *Natural Selection and Social Behavior: Recent Research and New Theory* (eds Alexander, R. D. & Tinkle, D. W.) 3–17 (Chiron, New York, 1981).
3. Fletcher, D. J. C. & Ross, K. G. *A. Rev. Entomol.* **30**, 319–343 (1985).
4. Strassmann, J. E. & Meyer, D. C. *Anim. Behav.* **31**, 431–438 (1983).
5. Sullivan, J. D. & Strassmann, J. E. *Behav. Ecol. Sociobiol.* **15**, 249–256 (1984).
6. Michener, C. D. in *Social Insects: An Evolutionary Approach to Castes and Reproduction* (ed. Engels, W.) 77–121 (Springer, Berlin, 1990).
7. Röseler, P.-F. in *The Social Biology of Wasps* (eds Ross, K. G. & Matthews, R. W.) 309–335 (Cornell University Press, Ithaca, 1991).
8. Fletcher, D. J. C. & Blum, M. S. *Science* **219**, 312–314 (1983).
9. Hölldobler, B. & Wilson, E. O. *The Ants* (Belknap, Cambridge, Massachusetts, 1990).
10. Velthuis, H. H. W., Ruttner, F. & Crewe, R. M. in *Social Insects: An Evolutionary Approach to Castes and Reproduction* (ed. Engels, W.) 231–243 (Springer, Berlin, 1990).

FIG. 1 *a*, Clockwise-rotating spiral wave in canine epicardial muscle. White, maximal depolarization; black, resting potential; numbers, time in ms. *b*, Time course of local activation on the upper left corner of the tissue during the last 10 cycles. Bar indicates time of recording in *a*. *c*, Snapshot of a spiral wave obtained in a computer simulation using the Fitz-Hugh-Nagumo model²³ in which anisotropy $D_x/D_y=4$ was introduced. Array size, 96×96 . Dashed box roughly corresponds to the size of the heart preparation shown in *a*.

METHODS. Slices ($20 \times 20 \times 0.5$ mm) of epicardial muscle were obtained from sheep and dogs (10–20 kg) and superfused with oxygenated Tyrode solution¹⁵ containing diacetyl monoxime (Sigma) 10–15 mM to abolish contractility^{15,26}. Tissues were stained with $1.3 \mu\text{g ml}^{-1}$ di-4- ΔNEPPS (Molecular Probes, Eugene, Oregon). Stimuli were delivered through one of four pairs of Ag–AgCl electrodes, as described elsewhere¹⁵. Self-sustaining repetitive activity was initiated as a result of the intersection of two perpendicular planar waves¹⁵. Refractory period was measured by the S_1 – S_2 protocol¹⁹. Fluorescence was excited at 490 nm and measured at 645 nm. Changes in fluorescence related to transmembrane potential changes^{14–17}, were recorded from an area of about 400 mm^2 , with a spatial resolution up to $30 \mu\text{m}$ per pixel, at a rate of 60 frames s^{-1} and 8 bits. Digital subtraction of background fluorescence and spatial filtering were



used. Equipment: Zenith 386 computer with a 361 Mbyte hard disk, an RGB colour monitor, a 4-Mbyte video board (Epix) and a video camera (Cohu series 6500). Hard copies were obtained with a video copy processor (Mitsubishi P40U).

s.d. 68 ms ; $n=33$) was 1.39 times longer than the refractory period ($T_r = 131 \pm 38 \text{ ms}$, $n=20$). A full revolution of a clockwise-rotating spiral and the pattern of local activation are shown in Fig. 1 *a*, *b*. The tip of the spiral has a pronounced curvature. As a result of the anisotropic properties of epicardial ventricular muscle¹⁹, wave propagation in the long axis of the cells was 3.9 ± 2.2 times faster than in the transverse direction, which resulted in an elliptical spiral. A similar elliptical spiral is observed in generic reaction-diffusion equations when an equivalent anisotropy is introduced (Fig. 1*c*).

Spiral waves are considered stationary when the tip of the spiral, which circulates around the core, follows a closed circular or elliptical trajectory and the periodicity is preserved everywhere outside the core²⁰. To characterize the size and location of the core we used a time-space plot ('frame stack') presentation. The time-space plots of the spiral wave (Fig. 2) have specific branching bands of lower amplitude in the centre of the plot which are divided in time by $T_s/2$. The distance between branching bands in the horizontal axis determines the effective size of the wave source (that is, the core) in this particular projection.

Figure 2*b* shows the frame-stack display of a recorded spiral wave. The rotating activity was stationary ($T_s = 115 \text{ ms}$) for more than 20 min (it was subsequently interrupted by applying external stimulation). The size of the core in the horizontal and vertical (not shown) projections was 1.1 mm and 9.7 mm, respectively. The size as well as the position of the core remained

stable throughout the episode. The real image of the preparation showed that the core was anchored to the border of a small (1×12 -mm) artery. In 14 episodes of stationary rotations analysed, the core was elongated (mean short and long axes, 3.1 ± 0.8 and $5.5 \pm 1.5 \text{ mm}$, respectively), and in nine cases showed similar attachment to anatomical discontinuities.

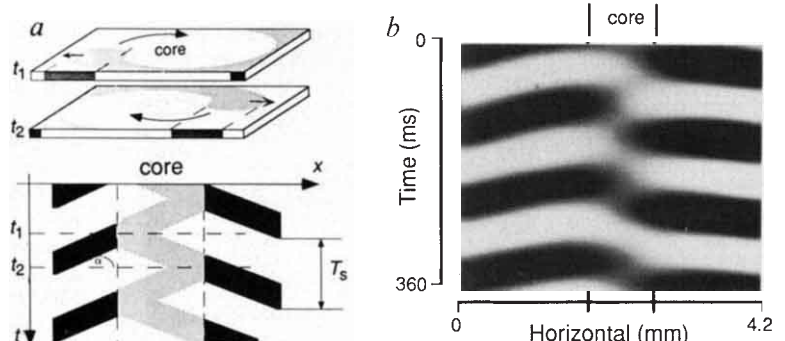
To determine whether the low-voltage activity in the core was solely produced by such discontinuities or was functionally determined, the voltage structure of the preparation during spiral wave activity (Fig. 3*a*) was compared with that during planar wave propagation (Fig. 3*b*). Frame-stack plots were obtained sequentially from narrow rectangular sections across the core's centre (Fig. 3*a*, top). The amplitude of fluorescence (depolarization) of 50 superimposed frames taken from the same section is shown in Fig. 3*a*, bottom. The amplitude was lower in the core's centre than in the periphery. In addition, the minimum value was higher, thus indicating that the core region never reached full repolarization. Signals of normal amplitude were recorded in the same region of tissue during planar wave propagation (Fig. 3*b*). Indeed, although the region contained a small artery, activity was uniform throughout the segment, which demonstrated that the differences in fluorescence associated with spiral wave activity are determined by functional differences between the core and the periphery. In 14 episodes of stationary spiral waves, the core amplitude was $25 \pm 17\%$ of the absolute maximum amplitude. Similar characteristics have been described for the BZ reaction²¹.

FIG. 2 *a*, Frame-stack display (similar to that in refs. 12, 27) of spiral wave activity obtained according to

$$\bar{F}(x, t) = \frac{1}{a} \int_0^a F(x, y, t) dy$$

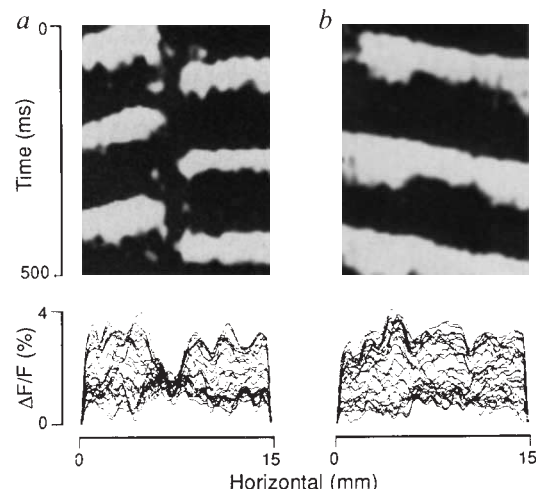
$$\bar{F}(y, t) = \frac{1}{b} \int_0^b F(x, y, t) dx$$

$F(x, y, t)$ is fluorescence signal, $a \times b$ is frame size. In each frame ($t_1, t_2, \dots, t_n$), all pixel values of each column were added to form a single line, represented at the edge of the frame. Arrows indicate direction of propagation. Intensity of shading indicates level of F ; tangent of α represents velocity of wave propagation. The $F(x, t)$ plot was obtained by stacking all 'frame-lines'. Black bands, excitation on either side of the core; T_s , vertical distance between two consecutive excitation bands; shaded bars, core. *b*, Frame-stack display (negative contrast) of a



stationary spiral wave induced in sheep epicardial muscle ($a \times b = 20 \times 26 \text{ mm}$). White regions, excitation bands; black regions, at rest. $T_s = 115 \text{ ms}$.

FIG. 3 *a*, Top: $F(x, t)$ plot of cross-section through centre of the core of a stationary spiral wave ($a \times b = 0.94 \times 15$ mm). Bottom: superimposed plots²² of $\Delta F/F$; 50 frames (roughly six rotating cycles) in the same region as on top. *b*, $F(x, t)$ and $\Delta F/F$ plots obtained 10 min later from the same section as in *a* during planar wave propagation.



Nonstationary rotations occur when the core shifts its position giving rise to spatial and temporal irregularities in the activation²⁰⁻²². Figure 4*a*, shows a frame-stack display obtained in one of seven episodes in which activity was nonstationary. The core initially appeared near the right border and slowly drifted 10 mm toward the left over a 1-s period (roughly seven cycles), in a direction that was perpendicular to the long axis of the cells, with no major changes in the vertical position (vertical frame-stack plot not shown). In general, however, the direction of the drift appeared to be independent of the cell orientation. In some cases the drift occurred in varying directions within the same preparation. Finally, in this experiment, the speed of wave front propagation in the direction of the drift was 90 mm s^{-1} . Overall, the average ratio of velocity of wave propagation over velocity of drift, was 9.8 ± 2.1 ($n = 7$).

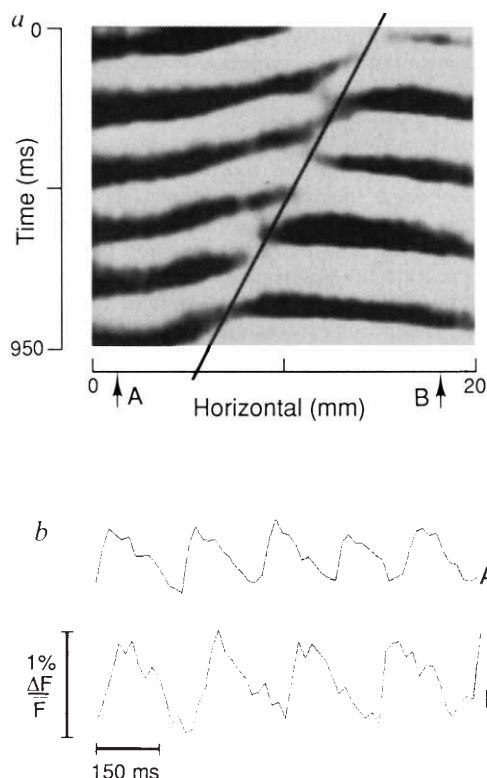


FIG. 4 *a*, $F(x, t)$ plot of a drifting spiral wave. Diagonal line follows the drift of the core in the horizontal axis. *b*, Pattern of local activation obtained from two points ahead and behind the core. The horizontal position of the selected points is indicated by the arrows in *a*. Mean T ahead (A) of the core was 175 ms; mean T behind (B) the core was 205 ms.

When the core of the drifting spiral collided with the border of the tissue, the spiral was annihilated and the rotations stopped. According to some theoretical suggestions²³, the spiral wave can interact with the boundary of the medium to result in drifting of the spiral along the boundary at a distance which is close to the critical radius, R_c (the minimum radius of that region which is necessary to maintain a spreading wave). In our experiments, however, the distance between the core and the border was never constant and there was no correlation between the border-to-core distance and drift velocity.

Drifting of the spiral wave resulted in a Doppler shift of the period of rotation (Fig. 4*b*). In the example shown, the rotation period ahead (A) of the core was 30 ms shorter than that behind (B), the core. In seven experiments, A (142 ± 27 ms) was always shorter than B (178 ± 35 ms). In some cases (Fig. 1), the drift direction varied, giving rise to aperiodic patterns of local activation (Fig. 1*b*). The analogy between spiral waves and re-entrant excitation which underlie life-threatening cardiac arrhythmias in experimental models^{13,18,24} as well as in the human heart²⁵ is well known^{1,2}. Our results strongly support this association. □

Received 8 October; accepted 14 November 1991.

1. Krinsky, V. I. (ed.) *Self-organization: Autowaves and Structures Far from Equilibrium* 9-18 (Springer, Berlin, 1984).
2. Winfree, A. T. *J. theor. Biol.* **138**, 353-405 (1989).
3. Keener, J. P. *J. appl. Math.* **46**, 1039-1056 (1986).
4. Gerhardt, M., Schuster, H. & Tyson, J. J. *Science* **247**, 1563-1566 (1990).
5. Markus, M. & Hess, B. *Nature* **347**, 56-58 (1990).
6. Wiener, N. & Rosenbluth, A. *Arch. Inst. Cardiol. Mex.* **16**, 205-265 (1946).
7. Krinsky, V. I. *Probl. Kibernetiki* **20**, 59-80 (1968).
8. Zhabotinsky, A. M. & Zaikin, A. N. *Oscillatory Processes in Biological and Chemical Systems* Vol. 2 279-283 (Pushchino, on Oka, 1971).
9. Winfree, A. T. *Science* **175**, 634-636 (1972).
10. Gerisch, G. *Curr. Topics dev. Biol.* **3**, 157- (1968).
11. Bures, J. & Gurelova, N. A. *J. Neurobiol.* **14**, 353-363 (1983).
12. Lechleiter, J., Girard, S., Peralta, E. & Clapham, D. *Science* **252**, 123-126 (1991).
13. Allesie, M. A., Bonke, F. I. M. & Shopman, F. J. *G. Circ. Res.* **33**, 54-62 (1973).
14. Salzberg, B. M., Davila, H. V. & Cohen, L. B. *Nature* **246**, 508-509 (1973).
15. Davidenko, J. M., Kent, P. F., Chialvo, D. R., Michaels, D. C. & Jalife, J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8785-8789 (1990).
16. Blaser, G. G. & Salama, G. *Nature* **321**, 579-585 (1986).
17. Kauer, J. *Nature* **331**, 166-168 (1988).
18. Frazier, D. W. *et al. J. clin. Invest.* **83**, 1039-1052 (1989).
19. Delgado, C., Steinhaus, B., Delmar, M., Chialvo, D. R. & Jalife, J. *Circ. Res.* **67**, 97-110 (1990).
20. Zykov, V. S. in *Simulation of Wave Processes in Excitable Media* (ed. Winfree, A. T.) 93-112 (1987).
21. Muller, S. C., Plesser, T. & Hess, B. *Science* **230**, 661-663 (1985).
22. Fast, V. G. & Pertsov, A. M. *Biophysics* **35**, 478-481 (1990).
23. Ermakova, E. A. & Pertsov, A. M. *Biophysics* **31**, 855-861 (1986).
24. Cardinal, R., Savard, P., Carson, L., Jean-Benoit, P. & Page, P. *Circulation* **70**, 136-148 (1984).
25. Downar, E. *et al. J. Am. Coll. Card.* **4**, 703-714 (1984).
26. Li, T., Sperelakis, N., Teneick, R. E. & Solaro, J. R. *J. Pharmac. exp. Ther.* **232**, 688-695 (1985).
27. Yamaguchi, T. & Mueller, S. C. *Physica D* **49**, 40-46 (1991).

ACKNOWLEDGEMENTS. We thank J. Kauer and G. Cinelli for help in developing our video optical mapping system, L. Lowe for voltage-sensitive dyes, A. Winfree, R. Ideker, V. Krinsky and D. Michaels for reading the manuscript and W. Coombs and J. A. Getchonis for technical assistance. This work was supported by the NIH and the American Heart Association.

Limbs generated at site of tail amputation in marbled balloon frog after vitamin A treatment

P. Mohanty-Hejmadi, S. K. Dutta & P. Mahapatra

Department of Zoology, Utkal University, Bhubaneswar 751004, Orissa, India

NIJAZI and Saxena¹ first observed that vitamin A has an inhibitory and modifying influence on tail regeneration in *Bufo andersonii* tadpoles. A positive relationship was later found between the inhibiting influence of vitamin A and the developmental stage of the regenerating tail in the same species². There have been several subsequent reports³⁻⁷ on the effects of vitamin A and its derivatives on limb development and regeneration. Thus in regenerating amphibian limbs, application of retinoids produces pattern duplication in the proximodistal and anteroposterior axes of the limb^{3,8,9}, and local application of retinoic acid to the anterior side of developing chick limbs causes duplications in the anteroposterior axis of limb^{10,11}. Here we show that vitamin A can cause limb development when applied to amputated tail stumps of the tadpoles of the marbled balloon frog *Uperodon systoma* (Anura Microhylidae). This is the first report of homeotic transformation mediated through vitamin A in vertebrates.

Following amputation through the middle of the tail at the hind-limb bud stage, tadpoles were exposed to a solution of 10 IU per ml vitamin A palmitate (Arovit, Roche; see Table 1 for details) for 24 h (set I), 48 h (set II), 72 h (set III), 96 h (set IV), 120 h (set V) and 144 h (set VI). Each set comprised 10 tadpoles which were reared according to standardized procedures¹². We found that vitamin A had a toxic effect on the development and survival of the tadpoles as 30–70% of the tadpoles died before the emergence of forelimbs. There was retardation in development as metamorphosis was delayed in the experimental sets compared with the control. Longer treatment suppressed normal growth of tails. The most unexpected result was the development of ectopic supernumerary limbs from the amputated tail stumps in all groups except set II, in which survival was the lowest. The percentage of tadpoles developing supernumerary limbs seems to be directly related to the period of exposure as it was 10% for 24 h exposure, increasing to 50% for 144 h. The number of supernumerary limbs in most cases was either two or three. Only one tadpole in set V (120 h) developed seven supernumerary limbs. These limbs had discernible thigh, shank, ankle and five digits, indicating that they were hind limbs. The number of limbs was always more than two, indicating that they developed in pairs initially, with a later duplication to account for the additional odd numbers occurring in sets V and VI (Fig. 1a–h). The histology and muscle pattern of these ectopic limbs is being investigated (P.M.-H., S.K.D. and P.M., unpublished results).

Staining with alcian blue and alizarin showed that the skeletal elements were cartilaginous; no ossification was observed. Normal limb development was partially affected as only 10 to 20% of the tadpoles developed abnormal limbs (Table 1). In set VI (144 h), in addition to the supernumerary limbs, an extra pair of hind limbs developed at the base of the original hind limb (Fig. 1h). Limb duplication has been observed before in regenerating amphibian limbs⁵, but never in developed limbs, although such duplication has been reported in developing limbs in the chick¹¹. The difference between regenerating and developing limbs may be that the wound associated with amputation is important for the induction of duplication⁵: in chick embryos duplications are only observed when vitamin A is administered by methods that cause a significant wound in the limb bud. In this study, limb duplication occurred after 144 h exposure to

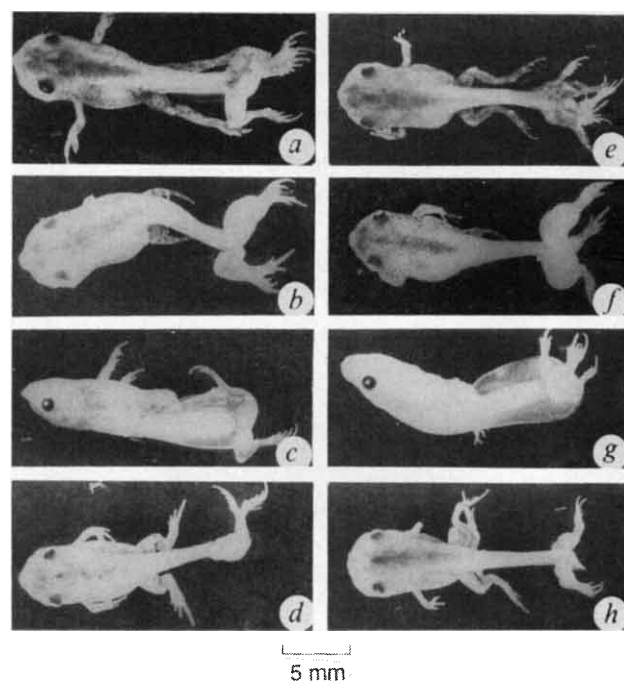


FIG. 1 Effects of vitamin A (10 IU) treatment for various times on limb generation from amputated tail stump. a, Treatment for 24 h, 3 limbs are generated; b, 72 h, 4 limbs; c and d, 96 h, 2 limbs; e, 120 h, 7 limbs; f, 120 h, 3 limbs; g, 144 h, 3 limbs; h, 144 h, 2 limbs, plus an extra pair of limbs below the original hindlimb.

vitamin A without inflicting any wound on the limb bud, which argues against this hypothesis⁵.

Our results show that vitamin A can cause homeotic transformation by inducing limb formation at ectopic sites. Similar results have been obtained by vitamin A treatment of the jumping frog, *Polypedates maculatus* (P.M.-H., P.M. and S.K.D., manuscript in preparation).

It is difficult to explain these startling results, but it is evident that in addition to the normal morphogenetic field for limb formation, other areas including the tail also have the capacity to organize limb formation. This indicates a broader morphogenetic field at the limb-bud stage, which probably becomes localized in the limb areas later on. So at the limb-bud stage the field for morphogenesis seems to be extendible and also inducible in neighbouring ectopic areas. This reveals a possible morphogenic action of vitamin A on the blastemal cells of the tail as well as the cells of the limb bud. Thus, a radical reprogramming of the cell genome can be induced by this vitamin. It is

TABLE 1 Effect of vitamin A on the tail stump after amputation in *Uperodon systoma* tadpoles

	0 h	Duration of vitamin A treatment*					
		24 h	48 h	72 h	96 h	120 h	144 h
Normal tail regeneration	10	1	0	0	0	0	0
Supernumerary limb from tail tip	0	1	0	2	2	3	5
Survival	10	6	3	5	7	5	6
Time (mean) for emergence of forelimbs (days)	13.6	18.4	15.6	15.4	24.8	27.6	24.8
Abnormality in hind limbs	0	0	1	2	0	1	1

* Vitamin A at 10 IU per ml. After tail amputation, tadpoles were raised in vitamin A solution for the required duration and then raised in conditioned water (up until the emergence of forelimb).

known that retinoic acid, a compound related to vitamin A, can activate homeobox genes (reviewed in ref. 13); our results may be evidence of a vitamin A-mediated induction of homeobox genes for limb formation. □

Received 26 July; accepted 4 November 1991.

1. Niazi, I. A. & Saxena, S. *Experientia* **24**, 852–853 (1968).
2. Niazi, I. A. & Saxena, S. *Indian J. exp. Biol.* **17**, 866–868 (1979).
3. Maden, M. *J. Embryol. exp. Morphol.* **77**, 273–295 (1983).
4. Maden, M., Keeble, S. & Cox, R. A. *Roux's Arch. dev. Biol.* **194**, 228–235 (1985).
5. Scadding, S. R. & Maden, M. *J. Embryol. exp. Morphol.* **91**, 19–34 (1986).
6. Scadding, S. R. & Maden, M. *J. Embryol. exp. Morphol.* **91**, 35–53 (1986).
7. Ludolph, D. C., Cameron, J. A. & Stocum, D. L. *Dev. Biol.* **140**, 41–52 (1990).
8. Thomas, S. D. & Stocum, D. *Dev. Biol.* **103**, 319–328 (1984).
9. Niazi, I. A. & Alam, S. *Roux's Arch. dev. Biol.* **193**, 111–116 (1984).
10. Tickle, C., Alberts, B., Walpert, L. & Lee, J. *Nature* **296**, 564–566 (1982).
11. Summerbell, D. *J. Embryol. exp. Morphol.* **78**, 269–289 (1983).
12. Mohanty-Hejmadi, P. *Prakriti-Utkal Univ. J. Sci.* **11**, 81–87 (1977).
13. De Robertis, E. M., Oliver, G. & Wright, C. V. E. *Scient. Am.* 46–52 (July, 1990).

ACKNOWLEDGEMENTS. P.M. thanks the University Grants Commission for a junior research fellowship.

Depletion of intracellular calcium stores activates a calcium current in mast cells

Markus Hoth & Reinhold Penner

Max-Planck-Institut für biophysikalische Chemie, Am Fassberg, 3400 Göttingen, Germany

IN many cell types, receptor-mediated Ca^{2+} release from internal stores is followed by Ca^{2+} influx across the plasma membrane^{1–3}. The sustained entry of Ca^{2+} is thought to result partly from the depletion of intracellular Ca^{2+} pools^{4,5}. Most investigations have characterized Ca^{2+} influx indirectly by measuring Ca^{2+} -activated currents^{6–9} or using Fura-2 quenching by Mn^{2+} , which in some cells enters the cells by the same influx pathway^{10,11}. But only a few studies have investigated this Ca^{2+} entry pathway more directly^{12–14}. We have combined patch-clamp and Fura-2 measurements to monitor membrane currents in mast cells under conditions where intracellular Ca^{2+} stores were emptied by either inositol 1,4,5-trisphosphate, ionomycin, or excess of the Ca^{2+} chelator EGTA. The depletion of Ca^{2+} pools by these independent mechanisms commonly induced activation of a sustained calcium inward current that was highly selective for Ca^{2+} ions over Ba^{2+} , Sr^{2+} and Mn^{2+} . This Ca^{2+} current, which we term I_{CRAC} (calcium release-activated calcium), is not voltage-activated and shows a characteristic inward rectification. It may be the mechanism by which electrically nonexcitable cells maintain raised intracellular Ca^{2+} concentrations and replenish their empty Ca^{2+} stores after receptor stimulation.

We have perfused inositol 1,4,5-trisphosphate ($\text{Ins}(1,4,5)\text{P}_3$) (10 μM) into mast cells through patch pipettes under conditions where Ca^{2+} currents are likely to be large, that is, clamping the cytosol to low intracellular Ca^{2+} concentrations ($[\text{Ca}^{2+}]_i$) using EGTA (10 mM) and increasing extracellular Ca^{2+} concentration ($[\text{Ca}^{2+}]_o$) (10 mM). Figure 1a illustrates that shortly after rupturing the patch that established continuity between patch pipette and cytosol (whole-cell configuration), a sustained inward current developed while the cell's membrane potential was clamped to 0 mV. The activation of the inward current by $\text{Ins}(1,4,5)\text{P}_3$ was observed in 111 out of 115 cells. The time course of the current was very similar in all the cells in which $\text{Ins}(1,4,5)\text{P}_3$ activated it, with delays following patch rupture of 7 ± 1 s (mean $\pm$ s.e.m., $n = 14$) and times from onset to half-maximal activation of 11 ± 1 s (mean $\pm$ s.e.m., $n = 12$). Because $\text{Ins}(1,4,5)\text{P}_3$ -induced Ca^{2+} release starts within 1 s of establishment of the whole-cell configuration^{12,15}, the delay in the appearance of the Ca^{2+} current indicates that $\text{Ins}(1,4,5)\text{P}_3$ is not likely

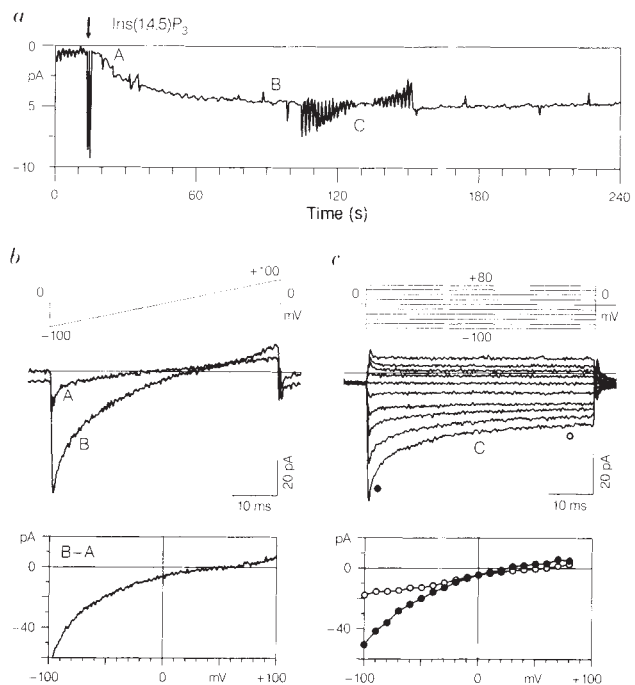


FIG. 1 Activation of the Ca^{2+} current by $\text{Ins}(1,4,5)\text{P}_3$. All data shown are taken from a single cell representative of a total of more than 100 similar experiments. **a**, Temporal pattern of activation of an inward current recorded at 0 mV holding potential during perfusion with the standard pipette solution supplemented with $\text{Ins}(1,4,5)\text{P}_3$ (10 μM) and EGTA (10 mM). Establishment of the whole-cell mode is indicated by the arrow. Data were sampled at 2 Hz. Immediately after breaking the patch, voltage ramps as shown in **b** were applied (occasionally sampled and seen as little spikes in the current trace). Labels A and B indicate times at which high-resolution current traces were taken for display in **b**, whereas C refers to traces obtained by square pulses shown in **c**. **b**, Single high-resolution current traces before (A) and after (B) development of the inward current. The voltage-ramp protocol is schematically shown on top. The lower graph illustrates the current-voltage relationship of the inward current as obtained by subtracting pooled leak-current ramps as in A from pooled steady-state current ramps as in B. **c**, Family of high-resolution current traces obtained after development of the inward current (C). The voltage-pulse protocol is schematically shown on top. The lower graph illustrates the current-voltage relationships taken at the times indicated by the circles (●, 0.5 ms; ○, 45 ms after the pulse). The data points were corrected by subtracting a linear leak current that was estimated from the slope of the ramp currents obtained before appearance of the inward current.

METHODS. Rat peritoneal mast cells were isolated and cultured as described²⁶. Patch-clamp experiments were done in the tight-seal whole-cell configuration at 23–27 °C in standard saline solution containing (in mM): NaCl, 140; KCl, 2.8; CaCl_2 , 10; MgCl_2 , 2; glucose, 11; HEPES-NaOH, 10, pH 7.2. Sylgard-coated patch pipettes had resistances between 2.5–5 M Ω after filling with the standard intracellular solution which contained (in mM): potassium-glutamate, 145; NaCl, 8; MgCl_2 , 1; Fura-2, 0.1; Mg-ATP, 0.5; EGTA, 10; HEPES-KOH, 10, pH 7.2. $\text{Ins}(1,4,5)\text{P}_3$ (10 μM , Amersham) and heparin (500 $\mu\text{g ml}^{-1}$; low M_r , from Sigma) were added to this solution when needed. Extracellular solution changes were made by local application from a wide-tipped micropipette. These pipettes contained the standard saline supplemented with ionomycin (10 $\mu\text{g ml}^{-1}$, Calbiochem) or, in the case of ion-substitution experiments, equimolar concentrations of BaCl_2 , SrCl_2 or MnCl_2 in exchange for 10 mM CaCl_2 . Possible contaminations of Ca^{2+} were chelated by the addition of EGTA (1 mM). The $[\text{Ca}^{2+}]_i$ was monitored with a photomultiplier-based system as described²⁷. The fluorescent indicator dye Fura-2 was loaded by diffusion from the patch pipette. $[\text{Ca}^{2+}]_i$ was calculated from the fluorescence ratio (360/390 nm) based on a calibration procedure as described²⁷. The two fluorescence intensities, DC current, and DC voltage were sampled at 2 Hz and filtered at 500 Hz. In addition, high-resolution current recordings were acquired by a computer-based patch-clamp amplifier system (EPC-9). Capacitive currents and series resistance were cancelled before each voltage ramp using the automatic capacitance compensation of the EPC-9. Currents were filtered at 2.3 kHz and digitized at 100 μs intervals. For analysis, currents were filtered digitally to 1 kHz.

to gate the current directly but rather secondarily, following depletion of intracellular Ca^{2+} stores.

To obtain current-voltage relationships for the current, we intermittently applied short (50 ms) voltage ramps from -100 to $+100$ mV (occasionally seen as spikes in the current trace of Fig. 1a). The very first ramps (labelled A in Fig. 1a) produced essentially a linear leak current (trace A in Fig. 1b). Subsequent ramps showed a gradually increasing inwardly rectifying current over almost the entire voltage range (trace B in Fig. 1b). The net current activated by $\text{Ins}(1,4,5)\text{P}_3$ in this experiment is obtained when subtracting trace A from trace B (current-voltage relationship in Fig. 1b). The rather positive reversal potential of the current suggests specificity for Ca^{2+} ions as charge carriers. The apparent reversal of the current at about $+50$ mV is probably underestimated, as most cells have chloride currents which develop under many experimental conditions^{12,16} and which often prevented an accurate assessment of the Ca^{2+} current at very positive potentials.

We tested whether the strong inward rectification at negative potentials is a genuine property of the inward current or caused by the ramp protocol. This was done by applying square voltage pulses over the same voltage range (Fig. 1c). The current responses to negative voltage pulses, where Ca^{2+} influx is largest, show some inactivation. Most of this inactivation is presumably Ca^{2+} -induced, as in preliminary experiments with the fast Ca^{2+} chelator BAPTA (1,2-bis(*O*-aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid), the inactivation was attenuated (data not shown). It thus seems that even 10 mM EGTA cannot buffer the incoming Ca^{2+} fast enough to prevent the inactivation. Both Ca^{2+} -induced inactivation and the ineffectiveness of EGTA in preventing it are also characteristic of voltage-activated Ca^{2+} channels¹⁷. The current-voltage relationships of the currents taken at the beginning and the end of each voltage pulse are depicted in Fig. 1c. When comparing these with the I - V curves obtained by voltage ramps, we conclude that the ramp protocol yields an appropriate representation of the current's behaviour, well aware of the fact that the ramp I - V curves just slightly exaggerate the curvature at negative potentials because of the Ca^{2+} -induced inactivation.

Although $\text{Ins}(1,4,5)\text{P}_3$ in conjunction with EGTA was the strongest activator of the Ca^{2+} current it seems unlikely that $\text{Ins}(1,4,5)\text{P}_3$ activated the current directly. This is corroborated

by the finding that simply omitting the $\text{Ins}(1,4,5)\text{P}_3$ from the pipette solution, but leaving the high concentration of EGTA (10 mM), often resulted in activation of a Ca^{2+} current that was qualitatively indistinguishable from the $\text{Ins}(1,4,5)\text{P}_3$ -induced conductance (Fig. 2a). This EGTA-induced activation was less reliable (21 out of 49 cells) and the delays as well as the times from onset to half-maximal activation were significantly longer compared with those of $\text{Ins}(1,4,5)\text{P}_3$ (mean delay after establishment of the whole-cell configuration: 44 ± 9 s, s.e.m., $n = 13$; mean time for half-maximal activation: 34 ± 9 s, s.e.m., $n = 10$). The range of responses is illustrated in Fig. 2a. EGTA could also activate I_{CRAC} in cells where large concentrations of heparin were added to the pipette solution (seven cells). Under these conditions, possible effects of $\text{Ins}(1,4,5)\text{P}_3$ should have been blocked¹⁸. We conclude the EGTA may empty intracellular Ca^{2+} stores by increasing the concentration gradients between cytosol and Ca^{2+} stores and prevents the refilling by acting as a scavenger. Under these conditions, it should mainly depend on how full and how 'leaky' the internal stores are, and whether, when and to what degree the influx pathway is activated.

Thapsigargin ($1 \mu\text{g ml}^{-1}$), a blocker of the Ca^{2+} pump of the endoplasmic reticulum¹⁹, could not induce activation of the Ca^{2+} current in cells where EGTA failed to do so (data not shown). This result is not unexpected, because the actions of thapsigargin should be equivalent to those of EGTA, namely preventing the reuptake of released Ca^{2+} into the stores. In the absence of intracellular EGTA, thapsigargin caused a rapid and sustained increase in $[\text{Ca}^{2+}]_i$ from basal levels of 312 ± 62 nM to $1,013 \pm 120$ nM (means $\pm$ s.e.m., $n = 6$). Part of this response was due to Ca^{2+} entry, probably through minimal activation of I_{CRAC} , which could not be resolved as its amplitude was less than 2 pA at -40 mV (data not shown). Again this is not unexpected, as under these unbuffered conditions the increase in $[\text{Ca}^{2+}]_i$ may prevent full activation of I_{CRAC} by reducing the driving force for both Ca^{2+} release and Ca^{2+} entry. In addition, elevated $[\text{Ca}^{2+}]_i$ may cause substantial inactivation of I_{CRAC} .

An efficient way of depleting intracellular stores is by applying ionomycin²⁰. Figure 2b illustrates an example of a cell in which EGTA did not activate the Ca^{2+} current and heparin was present to prevent a possible increase in $\text{Ins}(1,4,5)\text{P}_3$ formation. Subsequent application of ionomycin ($10 \mu\text{g ml}^{-1}$) induced an increase in $[\text{Ca}^{2+}]_i$ and the activation of I_{CRAC} , which in all

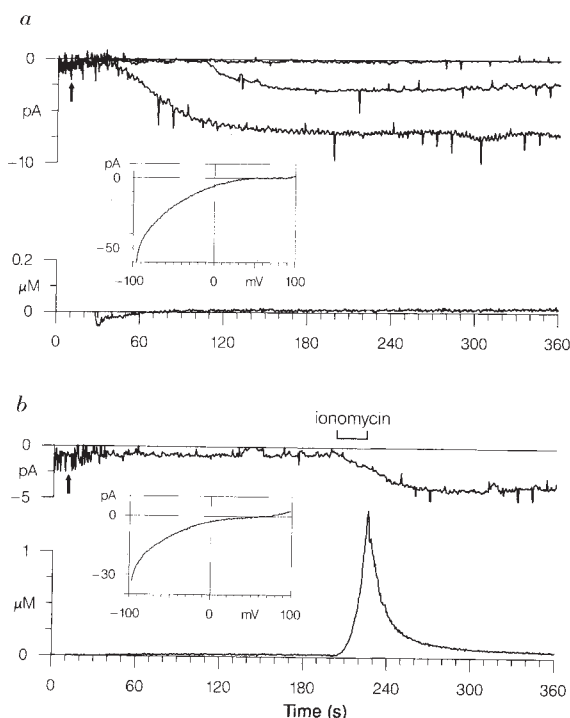


FIG. 2 Activation of the Ca^{2+} current by EGTA and ionomycin. *a*, Temporal pattern of activation of inward currents recorded at 0 mV holding potential during perfusion with the standard pipette solution which contained 10 mM EGTA. The superimposed current traces represent three different cells and demonstrate the range of observed responses. The bottom trace corresponds to $[\text{Ca}^{2+}]_i$ of the experiment with the largest current. In this experiment, which also showed the fastest time course of activation of all cells in this series, the pipette solution also contained heparin ($500 \mu\text{g ml}^{-1}$). The inset shows the current-voltage relationship of that current after full development (leak-corrected as described in Fig. 1b). *b*, Experimental conditions as in *a*. This cell did not develop an inward current in the presence of EGTA. It was, however, induced by application of ionomycin ($10 \mu\text{g ml}^{-1}$) at the time indicated.

cases strictly correlated with the administration of the ionophore (17 out of 17 cells). The $[Ca^{2+}]_i$ increase is probably due to translocation of $[Ca^{2+}]_o$ across the plasma membrane and is so large that it overwhelms the buffering capacity of EGTA. Ionomycin also releases Ca^{2+} from internal stores and this presumably activates the Ca^{2+} current. It is unlikely that the current is induced by ionomycin itself as the ionophore is thought to translocate Ca^{2+} by an electroneutral carrier mechanism²¹. Moreover, we have seen similar changes in $[Ca^{2+}]_i$, but have not detected additional current components associated with the application of ionomycin in cells in which the current was already fully activated by $Ins(1,4,5)P_3$ (12 out of 12 cells).

We assessed the specificity of this influx pathway by activating it with $Ins(1,4,5)P_3$ and EGTA-containing pipette solutions and have subsequently removed extracellular Ca^{2+} . This abolished the inward current (Fig. 3a), suggesting that the current is

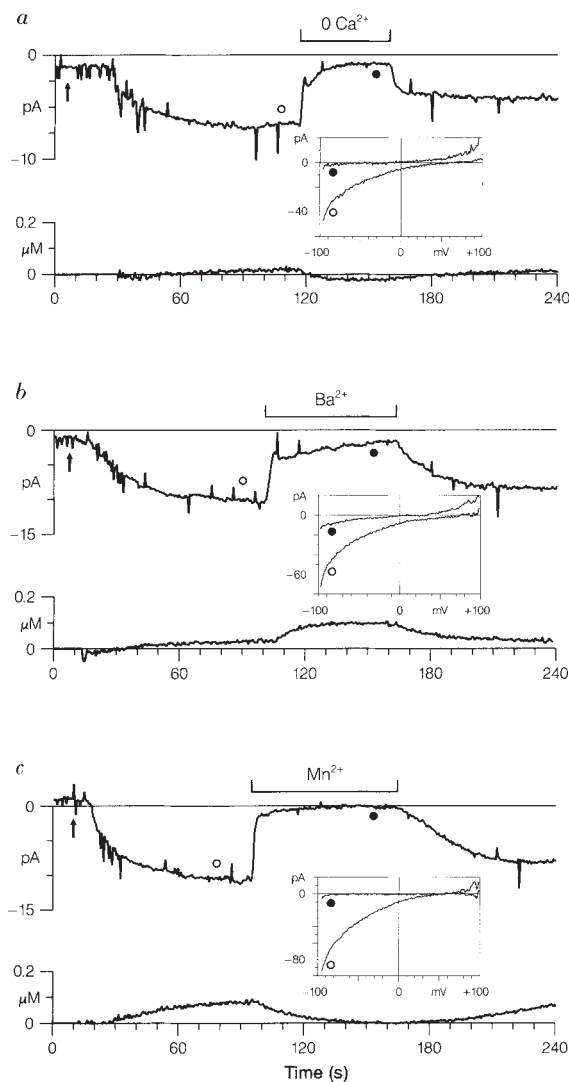


FIG. 3 Permeability of the inward current to divalent cations. In all three experiments the Ca^{2+} current was activated by pipette solutions containing $Ins(1,4,5)P_3$ ($10 \mu g ml^{-1}$) and EGTA ($10 mM$). The upper traces correspond to currents sampled at $0 mV$ and bottom traces show $[Ca^{2+}]_i$. The insets correspond to current-voltage relationships derived from subtracting pooled ramps shortly after break-in (before activation of the current) from pooled current responses at the times indicated by the respective symbols ($\circ$, inward current at $10 mM Ca^{2+}$; $\bullet$, current during application of test solution). a, standard saline ($10 mM Ca^{2+}$) was changed to Ca^{2+} -free saline ($1 mM EGTA$ added). b, Equimolar substitution of $10 mM Ca^{2+}$ in standard saline by $10 mM Ba^{2+}$ ($1 mM EGTA$ added). c, Equimolar substitution of $10 mM Ca^{2+}$ in standard saline by $10 mM Mn^{2+}$ ($1 mM EGTA$ added).

specific for divalent ions. Most Ca^{2+} channels known are highly permeant to Ba^{2+} and Sr^{2+} . But I_{CRAC} seems not to be carried by these ions. Figure 3b shows that the equimolar substitution of extracellular Ca^{2+} by Ba^{2+} and addition of $1 mM EGTA$ clearly reduces the inward current. This effect of Ba^{2+} was observed in five out of five cells and similar results were obtained with Sr^{2+} (two out of two cells). It is not yet clear whether Ba^{2+} is impermeable or the Ca^{2+} influx pathway requires Ca^{2+} at the extracellular side to function. Another possible explanation for the reduction in Ca^{2+} current could be an inactivation of the current from the inside by an initial entry of Ba^{2+} into the cell. In Fig. 3b there is an increase in the fluorescence-ratio trace which may be attributed to Ba^{2+} entering the cytosol and affecting Fura-2 fluorescence. We noticed that the increase in the ratio was due to an increase in the fluorescence at $360 nm$ excitation (isostilbic point of Fura-2 for Ca^{2+}) and a decrease in fluorescence at $390 nm$ excitation (data not shown). This suggests that the isostilbic point of Fura-2 with respect to Ba^{2+} is shifted to longer wavelengths and may be used to monitor Ba^{2+} fluxes in cells where $[Ca^{2+}]_i$ is strongly buffered by EGTA.

In some cell types Mn^{2+} seems to permeate through the same Ca^{2+} influx pathway as Ca^{2+} does^{10,22-24} (but see ref. 25). With as little as $50 \mu M$ of extracellular Mn^{2+} , its entry can be measured by a large quench of Fura-2 fluorescence in the cytosol¹¹. We have therefore substituted extracellular Ca^{2+} by $10 mM Mn^{2+}$ and added $1 mM EGTA$. When applying this solution to a cell that had developed a Ca^{2+} current, there was a strong block of the current and a concomitant decrease in $[Ca^{2+}]_i$ (Fig. 3c). However, there was only a very small decrease in the fluorescence signal obtained at $360 nm$ excitation, suggesting little if any Mn^{2+} entry through the influx pathway. This would suggest that in other cell types Mn^{2+} can enter through other pathways (for example receptor-operated channels or ion transporters), or that Mn^{2+} entry occurs through a similar current as described here with a different selectivity for divalent ions.

We have shown that depleting intracellular Ca^{2+} stores by three independent mechanisms activates a highly selective inwardly rectifying Ca^{2+} current in mast cells. As ionomycin and EGTA can activate I_{CRAC} in the presence of heparin, we do not find that inositol phosphates ($Ins(1,4,5)P_3$ or $Ins(1,3,4,5)P_4$) are necessary for activating calcium influx (as proposed in ref. 2). The characteristics of this current are different from classic voltage-activated Ca^{2+} currents in that Ba^{2+} and Sr^{2+} do not seem to permeate this pathway readily. At present, we do not know how the emptying of the intracellular Ca^{2+} stores translates into activation of a plasma membrane current. This might be due to a second messenger released from the storage organelles or a direct coupling between proteins in the two membranes. □

Received 12 August; accepted 4 October 1991.

- Berridge, M. J. & Irvine, R. F. *Nature* **341**, 197-205 (1989).
- Irvine, R. F. *FEBS Lett.* **263**, 5-9 (1990).
- Meldolesi, J., Clementi, E., Fasolato, C., Zacchetti, D. & Pozzan, T. *Trends pharmacol. Sci.* **12**, 289-292 (1991).
- Castells, R. & Droogmans, G. *J. Physiol., Lond.* **317**, 263-279 (1981).
- Putney, J. W. Jr *Cell Calcium* **11**, 611-624 (1990).
- Bird, G. St J. *et al. Nature* **352**, 162-165 (1991).
- Morris, A. P., Gallacher, D. V., Irvine, R. F. & Petersen, O. H. *Nature* **330**, 653-655 (1987).
- Liano, I., Marty, A. & Tanguy, J. *Pflügers Arch.* **409**, 499-506 (1987).
- Snyder, P. M., Krause, K.-H. & Welsh, M. J. *J. Biol. Chem.* **263**, 11048-11051 (1988).
- Sage, S. O., Merritt, J. E., Hallam, T. J. & Rink, T. J. *Biochem. J.* **258**, 923-926 (1989).
- Jacob, R. *J. Physiol., Lond.* **421**, 55-77 (1990).
- Penner, R., Matthews, G. & Neher, E. *Nature* **334**, 499-504 (1988).
- Matthews, G., Neher, E. & Penner, R. *J. Physiol., Lond.* **418**, 105-130 (1989).
- Lewis, R. S. & Cahalan, M. D. *Cell Regulation* **1**, 99-112 (1989).
- Marty, A. & Tan, Y. P. *J. Physiol., Lond.* **419**, 665-687 (1989).
- Matthews, G., Neher, E. & Penner, R. *J. Physiol., Lond.* **418**, 131-144 (1989).
- Eckert, R. & Chad, J. E. *Prog. Biophys. molec. Biol.* **44**, 215-267 (1984).
- Cullen, P. J., Comerford, J. G. & Dawson, A. P. *FEBS Lett.* **228**, 57-59 (1988).
- Thastrup, O., Cullen, P. J., Drobak, B. K., Hanley, M. R. & Dawson, A. P. *Proc. natn. Acad. Sci. USA* **87**, 2466-2470 (1990).
- Albert, P. R. & Tashjian, A. H. Jr *J. Biol. Chem.* **259**, 15350-15363 (1984).
- Liu, C. & Hermann, T. E. *J. Biol. Chem.* **253**, 5892-5894 (1978).
- Merritt, J. E., Jacob, R. & Hallam, T. J. *J. Biol. Chem.* **253**, 1522-1527 (1989).
- Mertz, L. M., Baum, B. J. & Ambudkar, I. S. *J. Biol. Chem.* **265**, 15010-15014 (1990).

24. Kass, G. E. N., Llopis, J., Chow, S. C., Duddy, S. K. & Orrenius, S. *J. Biol. Chem.* **265**, 17486–17492 (1990).
 25. Llopis, J., Chow, S. B., Kass, G. E. N., Gahm, A. & Orrenius, S. *Biochem. J.* **277**, 553–556 (1991).
 26. von zur Mühlen, F., Eckstein, F. & Penner, R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 926–930 (1991).
 27. Neher, E. in *Neuromuscular Junction* (eds Sellin, L. C., Libelius, R. & Thesleff, S.) 65–76 (Elsevier, Amsterdam, 1989).

ACKNOWLEDGEMENTS. We thank E. Neher for critical comments on the manuscript. This work was supported in part by the Deutsche Forschungsgemeinschaft and the Sonderforschungsbereich 236.

Inositol 1,3,4,5-tetrakisphosphate activates an endothelial Ca^{2+} -permeable channel

Andreas Lückhoff & David E. Clapham

Department of Pharmacology, Mayo Foundation, Rochester, Minnesota 55905, USA

RECEPTOR-MEDIATED increases in the cytosolic free calcium ion concentration in most mammalian cells result from mobilization of Ca^{2+} from intracellular stores as well as transmembrane Ca^{2+} influx. Inositol 1,4,5-trisphosphate (InsP_3) releases calcium from intracellular stores¹ by opening a Ca^{2+} -permeable channel in the endoplasmic reticulum^{2–4}. But the mechanism and regulation of Ca^{2+} entry into nonexcitable cells has remained elusive because the entry pathway has not been defined. Here we characterize a novel inositol 1,3,4,5-tetrakisphosphate (InsP_4) and Ca^{2+} -sensitive Ca^{2+} -permeable channel in endothelial cells. We find that InsP_4 , which induces Ca^{2+} influx into acinar cells^{5,6}, enhances the activity of the Ca^{2+} -permeable channel when exposed to the intracellular surface of endothelial cell inside-out patches. Our results suggest a molecular mechanism which is likely to be important for receptor-mediated Ca^{2+} entry.

Manganese ions may use the same entry pathway as Ca^{2+} in stimulated endothelial cells^{7,8}. To quantify Mn^{2+} influx, endothelial cells patch-clamped with the nystatin technique⁹ and stimulated with ATP (30 μM) were exposed to Mn^{2+} (6 mM). Figure 1 shows that at negative membrane potentials, close to the potassium equilibrium potential (E_K), an inward current of 40 ± 24 (s.d., $n = 8$) pA (3.2 ± 2.5 pA pF⁻¹) was activated by ATP. No such Mn^{2+} current was observed in unstimulated cells (not shown) or in depolarized cells, in accordance with previous reports that endothelial cation influx is inhibited by depolarization^{10–12}.

Using patch-clamp techniques¹³, Mn^{2+} -permeant single channels were identified in cell-attached, inside-out and outside-out patches. For these experiments, Mn^{2+} (110 or 140 mM) was used as the dominant charge carrier. Channel events observed in inside-out patches are shown in Fig. 2a. Channel activity was observed at membrane potentials ranging from -160 to -10 mV, with a slope conductance of 2.5 pS and a mean open time of about 230 ms at -80 mV. The low single-channel conductance prevented rigorous analysis of the channel kinetics. Glutamate replacement of chloride in the bath did not affect channel conductance. When Ca^{2+} or Ba^{2+} replaced Mn^{2+} in the pipette, no measurable change in the conductance from -120 to -60 mV was noted, suggesting that the channel is roughly equally permeable to these divalent cations. The channel was more selective for divalent than sodium ions ($>2:1$) as shown in outside-out patch experiments ($n = 3$; Fig. 2d) but determination of absolute permeability ratios is prohibited by the low channel conductance in realizable ionic concentrations.

Typically, inside-out patches exhibited a high channel activity immediately after excision and a considerable run-down over 2 to 6 min. The channel was sensitive to internal Ca^{2+} concentrations; addition of EDTA or EGTA to the bath, resulting in Ca^{2+} concentrations below 10 μM , gradually decreased

the open probability of the channel (Fig. 2c) that was restored on readmission of Ca^{2+} .

Remarkably, addition of InsP_4 to inside-out patches displaying channel openings on activation by high Ca^{2+} concentrations (1 mM), evoked an immediate increase in the open probability (Fig. 3). InsP_4 (10–100 μM) increased channel activity (defined as the product of the number of channels, N , and the probability of channel opening, P_o) by a factor of 2 to 8 (mean 3.9 ± 2.0 ; $n = 12$). Activation by InsP_4 was reversible on washout (Fig. 3), but rundown of channel activity prevented intrapatch definition of the InsP_4 concentration-effect relation. InsP_4 (10–30 μM) increased channel activity at lower Ca^{2+} concentrations (0.47 μM) by a factor of 5.0 (± 2.8 ; $n = 7$). No significant activation was induced by InsP_4 at a Ca^{2+} concentration of 0.1 μM . The limited analysis possible of channel kinetics suggested that the effects of InsP_4 were due to an increased frequency of openings rather than a prolongation of open time. In summary, basal channel activity scales with Ca^{2+} concentration and a minimum level of intracellular Ca^{2+} seems to be necessary for InsP_4 -dependent channel activation.

In contrast to InsP_4 , InsP_3 (50 and 100 μM) failed to induce any obvious increase in channel activity in 22 patches. In 10 patches, InsP_4 , applied after InsP_3 stimulated the channel (Fig. 4). There was no indication that the effect of InsP_4 was enhanced by pretreatment with InsP_3 (Fig. 4c). 1,3,4,5,6- InsP_5 (500 μM ; $n = 3$) enhanced channel activity at concentrations 50-fold higher than InsP_4 but not at 100 μM . Furthermore, GTP- γ -S (10 μM , in bath or pipette solution), alkalization of the bath (pH 7.8), stretch (application of negative pressure up to -25 cm H₂O), and depolarization from -90 to -30 mV, all failed to activate the channel. Channel activity was blocked reversibly by bath application of 1–5 mM Ni^{2+} .

Although channel events were observed in excised patches in about one-third of experiments, channels were rarely detected in the cell-attached configuration in unstimulated cells. But

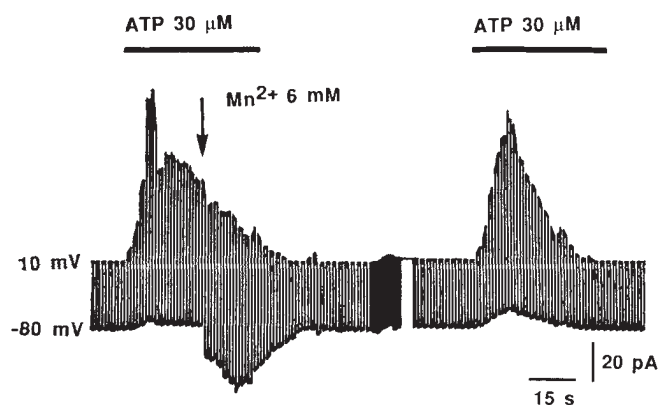


FIG. 1 Mn^{2+} influx into stimulated endothelial cells. An endothelial cell under perforated-patch whole-cell voltage clamp was alternately stepped from $+10$ to -80 mV each second. The cell was superfused with ATP (30 μM , bar), inducing a large outward current at $+10$ mV (largely calcium-activated K^+ current). During the first stimulation, 6 mM Mn^{2+} was added to the ATP solution perfusing the cell (for ~ 5 s from arrow), inducing an inward current at -80 mV. A recovery time of 5 min was allowed during the two stimulations. METHODS. Cultured endothelial cells from bovine aortae (subculture 13–30)²⁵ were grown on glass coverslips to subconfluence in Dulbecco's modified Eagle's medium with 20% fetal calf serum. The perforated-patch whole-cell voltage clamp technique²⁶ was done using nystatin as a pore-forming agent as described⁹. The patch-clamp pipette contained (in mM) KCl (40), K_2SO_4 (80), MgCl_2 (2), HEPES (15), pH 7.4, nystatin 0.1–0.3 mg ml⁻¹. Standard bath solution (SBS) contained NaCl (140), KCl (4), CaCl_2 (1), MgCl_2 (2), HEPES (10), glucose (5), pH 7.4. Superfusion with ATP solution (30 μM in CaCl_2 -free SBS) was at 0.2 ml min⁻¹ using a U-shaped glass capillary tube with an outlet hole placed next to the cell. Mn^{2+} was applied by introducing CaCl_2 -free SBS with 30 μM ATP and 6 mM Mn^{2+} into a small tube inside the U-tube, with its outlet adjacent to the mouth of the U-tube. Records were filtered at 100 Hz. All experiments were done at $22 \pm 2^\circ\text{C}$.

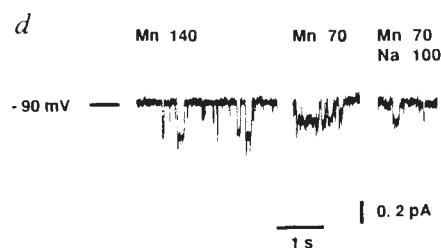
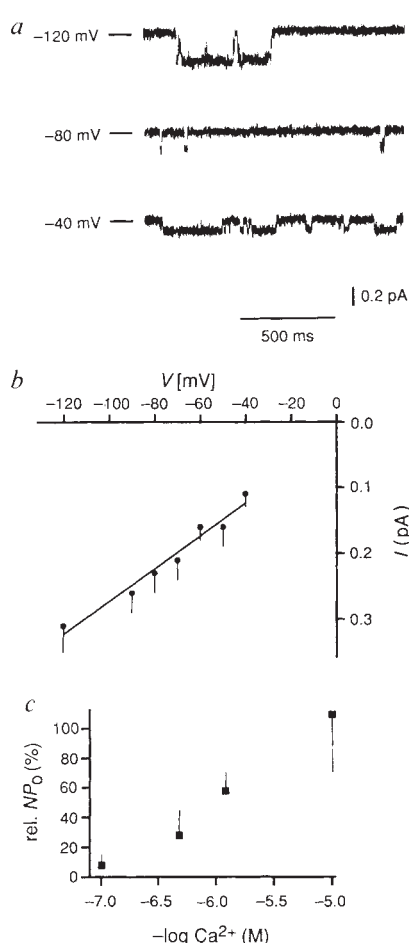


FIG. 2 Single Mn^{2+} -channel activity recorded from inside-out and outside-out patches excised from endothelial cells. *a*, Sample tracings from one inside-out patch at three holding potentials; *b*, Single-channel amplitudes plotted as a function of holding potential (mean $\pm$ s.d.; $n=15$). The regression line yields a slope conductance of 2.5 pS. *c*, Channel activity (NP_0) of inside-out patches (-70 to -90 mV) plotted as a function of bath Ca^{2+} concentration. Channel activity was normalized to average NP_0 recorded at 1 mM Ca^{2+} bath concentration. *d*, Sample tracings from an outside-out patch with different cation concentrations, as shown, in the bath solution. Replacement of 50% of bath divalents by equimolar choline (100 mM) reduced the channel amplitude by $\sim 40\%$. Subsequent replacement of choline by Na^+ did not increase channel amplitude. Assuming the channel permeability obeys the Goldman-Hodgkin-Katz relation and that we could detect a 15% change in mean channel amplitude, the channel is at least twice as permeable to Mn^{2+} as it is to Na^+ .

METHODS. Patches were excised from confluent cultured endothelial cells. For inside-out patches¹³, the bath contained SBS (see Fig. 1 legend) and the pipette solution contained (in mM): $MnCl_2$ (110), $BaCl_2$ (6), $CaCl_2$ (1), $MgCl_2$ (2), HEPES (5), pH 7.1. Changes in the Ba^{2+} concentration (used to block inward rectifying K^+ channels) from 3 to 12 mM did not noticeably change the current amplitudes. For outside-out patches¹³, the bath contained $MnCl_2$ (140), $BaCl_2$ (10), $CaCl_2$ (1), $MgCl_2$ (2), HEPES (10), pH 7.2. $MnCl_2$ was reduced to 70 mM, $BaCl_2$ to 5 mM, by substitution with either choline chloride (100 mM) or NaCl (100 mM) (maintaining osmolality). The pipette contained choline chloride (140), $CaCl_2$ (1), $MgCl_2$ (2), HEPES (10), pH 7.4. The pipette tip was filled with solution containing 2 mM EGTA to establish the outside-out configuration. Data were filtered at 3 kHz, digitized at 8 kHz, and stored on videotape. The tracings of Figs 2a, 3 and 4 are filtered at 400 Hz (eight-pole Bessel filter) and those of Fig. 2d at 100 Hz. The analysed channel kinetics and open probabilities were done with data filtered at 100 Hz. Mean open times were calculated from a single-exponential fit of events at -60 to -90 mV plotted in logarithmic bins²⁷. NP_0 was calculated as described²⁸.

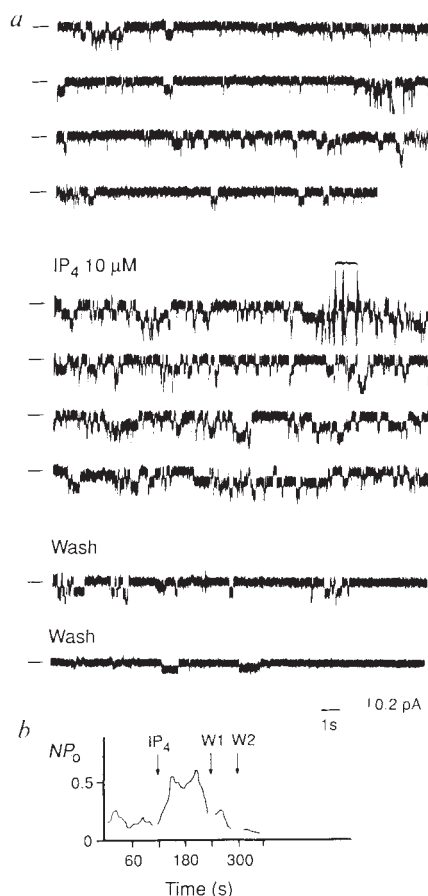


FIG. 3 Channel activation by $InsP_4$. *a*, $10 \mu M$ $InsP_4$ (purity $>99\%$) in bath solution was applied to the inside surface of an excised inside-out patch after a control period (top). During each washout, $\sim 50\%$ of the bath volume was exchanged. Recordings during addition of $InsP_4$ (over about 15 s) are not shown but the tracings are otherwise continuous until wash 1. The two bottom traces (wash) show the first 30 s after each washout. The bracket marks one electrical artefact. *b*, Calculated single-channel activity (NP_0 ; plotted as running average over 20 s) shown continually during the experiment in *a*. $InsP_4$ reversibly increased channel activity 2.5-fold over control in this patch.

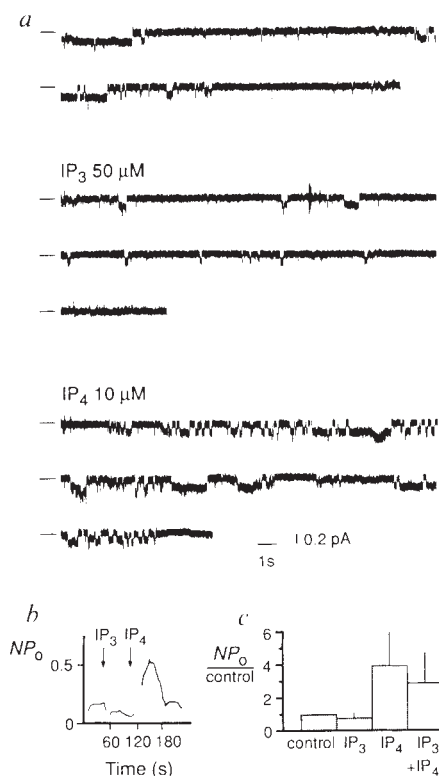


FIG. 4 InsP₃ does not induce nor potentiate activation of InsP₄-sensitive channels. *a*, 50 μM InsP₃ (Calbiochem) and InsP₄ (10 μM) were consecutively added to an endothelial inside-out patch held at -80 mV. The tracings are continuous except during addition of compound. *b*, Calculated channel activity of the experiment shown in *a*. *c*, Summary of channel activity after InsP₄ (10–30 μM) and InsP₃ (50 μM) applications to endothelial patches in 1 mM Ca²⁺, shown relative to controls. InsP₄ consistently increased channel activity in patches where InsP₃ had no effect. Addition of InsP₃ to patches activated by InsP₄ did not potentiate channel activity. InsP₃ was tested for potency by injection into *Xenopus laevis* oocytes, where it released intracellular calcium as measured under confocal microscopy²⁹.

channel openings in cell-attached patches (intact cells) were induced by addition of ATP (30 μM; $n = 4$ out of 9 experiments in which channel activity was seen after excision), bradykinin (30 nM; $n = 2/7$), and inositol (0.3 or 1 mM; $n = 16/25$) to the bath. None of these substances stimulated the channel when added to the pipette solution in cell-attached patches or to the bath for excised patches. In a limited number of experiments in which Ca²⁺ (110 mM) replaced Mn²⁺ in the pipette (calcium was often associated with vesicle formation, limiting the number of these experiments), channel activity was induced by inositol (1 mM) in cell-attached patches and was increased by InsP₄ (30 μM) in Ca²⁺-activated inside-out patches ($n = 3$).

The Ca²⁺ and InsP₄-sensitive channel reported here is clearly different from other channels suggested to mediate Ca²⁺ entry in endothelial cells (for review, see refs 14, 15). We observed an ~15 pS sodium-permeable channel from 0 to -40 mV in inside-out patches when 140 mM Na⁺ was in the pipette. This channel, unlike the Mn²⁺-permeable channel, was insensitive to InsP₄, did not run down rapidly, was not rapidly blocked by Ni²⁺, and had long open times. There are several lines of evidence to suggest that the Mn²⁺ channel described here accounts for previously described receptor-mediated cation entry into endothelial cells. First, the channel is permeable to Ca²⁺ as well as to Mn²⁺ (see refs 7, 8). The channel is activated by receptors coupled to endothelial inositol phosphate turnover and subsequent calcium influx in intact cells, such as ATP and bradykinin^{16–19}. Finally, the current is decreased by depolarization^{10–12} and is inhibited^{7,8} by Ni²⁺ (albeit internally in this study). Although the fact that the channel requires Ca²⁺ for activation in the inside-out configuration cannot be reconciled with the finding that Mn²⁺ influx may occur in cells with very low cytosolic Ca²⁺ concentrations⁸, Ca²⁺ influx is normally associated with InsP₃-induced Ca²⁺ mobilization. Under these conditions, Ca²⁺ and InsP₄ may cooperatively activate the channel. In fact, the data suggest that high intracellular Ca²⁺, as occurs after inositol phosphate activation, permits InsP₄ gating of the channel, whereas low intracellular Ca²⁺ concentrations would prohibit activation by InsP₄. Agonist-induced oscillations of cytosolic free Ca²⁺ concentration ([Ca²⁺]_c)²⁰ may be gener-

ated by this, as well as by other forms of Ca²⁺-induced increases in [Ca²⁺]_c (refs 4, 21).

One may speculate that the endothelial Ca²⁺ channel is present in other nonexcitable cells, supporting Irvine's hypothesis²² for an InsP₄ role in gating calcium entry to replenish intracellular stores. This would not contradict reports that Ca²⁺ influx may occur independently²³ or even in the absence²⁴ of InsP₄ because the channel we describe was active in the absence of any stimulus other than Ca²⁺. But our results may provide the molecular basis for the profound activation of Ca²⁺ influx by InsP₄ under some experimental conditions. □

Received 27 September; accepted 2 December 1991.

1. Streb, H., Irvine, R. F., Berridge, M. J. & Schulz, I. *Nature* **306**, 67–68 (1983).
2. Ferris, C. D., Hagan, R. L., Supattapone, S. & Snyder, S. H. *Nature* **342**, 87–89 (1989).
3. Furukawa, T. et al. *Nature* **342**, 32–38 (1989).
4. Bezprozvanny, I., Watras, J. & Ehrlich, B. E. *Nature* **351**, 751–754 (1991).
5. Changya, L., Gallacher, D. V., Irvine, R. F. & Petersen, O. H. *J. Membrane Biol.* **109**, 85–93 (1989).
6. Morris, A. P., Gallacher, D. V., Irvine, R. F., Potter, B. V. L. & Petersen, O. H. *Nature* **330**, 653–655 (1987).
7. Hallam, T. J., Jacob, R. & Merritt, J. E. *Biochem. J.* **255**, 179–184 (1988).
8. Jacob, R. *J. Physiol., Lond.* **421**, 55–77 (1990).
9. Horn, R. & Marty, A. *J. gen. Physiol.* **92**, 145–159 (1988).
10. Lückhoff, A. & Busse, R. *Pflügers Arch.* **416**, 305–311 (1990).
11. Schilling, W. P. *Am. J. Physiol.* **257**, H778–H784 (1989).
12. Schilling, W. P., Rajan, L. & Strobl-Jager, E. *J. Biol. Chem.* **264**, 12838–12848 (1989).
13. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch.* **391**, 85–100 (1981).
14. Adams, D. J., Barak, J., Laskey, R. & van Breemen, C. *FASEB J.* **3**, 2389–2400 (1989).
15. Nilius, B. *News Physiol. Sci.* **6**, 110–114 (1991).
16. Colden-Stanfield, M. et al. *Circ. Res.* **61**, 632–640 (1987).
17. Derian, C. K. & Moskowitz, M. A. *J. Biol. Chem.* **261**, 3831–3837 (1986).
18. Lückhoff, A. & Busse, R. *J. cell. Physiol.* **126**, 414–420 (1986).
19. Pirotton, S., Raspe, E., Demolle, D., Erneux, C. & Boeynaems, J. M. *J. Biol. Chem.* **262**, 17461–17466 (1987).
20. Jacob, R., Merritt, J. E., Hallam, T. J. & Rink, T. J. *Nature* **335**, 40–45 (1988).
21. Finch, E. A. F., Turner, T. J. & Goldin, S. M. *Science* **252**, 443–446 (1991).
22. Irvine, R. F. *FEBS Lett.* **263**, 5–9 (1990).
23. Penner, R., Matthews, G. & Neher, E. *Nature* **334**, 499–504 (1988).
24. Bird, G. S. J. et al. *Nature* **352**, 162–165 (1991).
25. Schwartz, S. M. *In Vitro* **14**, 966–980 (1978).
26. Lindau, M. & Fernandez, J. *Nature* **319**, 150–153 (1986).
27. Sigworth, F. J. & Sine, S. M. *Biophys. J.* **52**, 1047–1054 (1987).
28. Logothetis, D. E., Kurachi, Y., Galper, J., Neer, E. J. & Clapham, D. E. *Nature* **325**, 321–326 (1987).
29. Leichter, J., Girard, S., Peralta, E. & Clapham, D. E. *Science* **252**, 123–126 (1991).

ACKNOWLEDGEMENTS. We thank R. Irvine for InsP₄ and InsP₃, P. Davies for bovine aorta endothelial cells, S. Manahan for tissue culture and M. Lückhoff for help with the figures. The study was supported by NIH grants to D.E.C. and a Heisenberg Fellowship of the Deutsche Forschungsgemeinschaft to A.L.D.E.C. is an Established Investigator of the American Heart Association.

9-*Cis* retinoic acid stereoisomer binds and activates the nuclear receptor RXR α

Arthur A. Levin, Laurie J. Sturzenbecker, Sonja Kazmer, Thomas Bosakowski, Christine Huselton*, Gary Allenby, Jeffrey Speck, Cl. Kratzenisen, Michael Rosenberger†, Allen Lovey† & Joseph F. Grippo

Departments of Toxicology and Pathology, † Medicinal Chemistry and * Drug Metabolism, Hoffmann-La Roche, Nutley, New Jersey 07110, USA

VITAMIN A (retinol) and its natural derivatives are required for many physiological processes¹⁻³. The activity of retinoids is thought to be mediated by interactions with two subfamilies of nuclear retinoic acid receptors, RAR and RXR. The RARs bind all-*trans* retinoic acid (*t*-RA) with high affinity and alter gene expression as a consequence of this direct ligand interaction⁴⁻¹⁰. RXR α is activated by *t*-RA, yet has little binding affinity for this ligand (ref. 11). *t*-RA may be converted to a more proximate ligand that directly binds and activates RXR α , and we have developed a method of nuclear receptor-dependent ligand trapping to test this hypothesis. Here we report the identification of a stereoisomer of retinoic acid, 9-*cis* retinoic acid, which directly binds and activates RXR α . These results suggest a new role for isomerization in the physiology of natural retinoids.

We developed a method to trap derivatives of *t*-RA bound to RXR α in COS-1 cells. Transient transfection of COS cells with expression vectors for RAR α , β or γ yields nucleosol fractions highly enriched for [³H]*t*-RA binding associated with these receptors¹²⁻¹⁵. We reasoned that COS-1 cell nuclei enriched with RXR α may selectively bind ligands that have

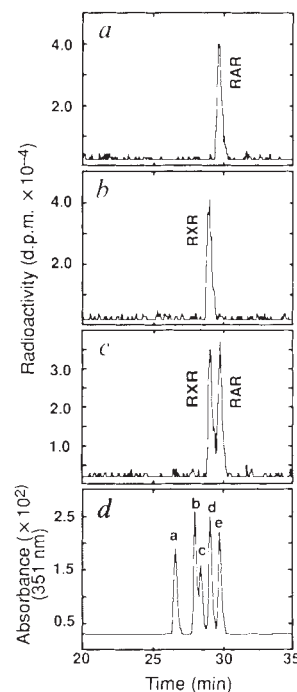
high affinity for RXR α . If *t*-RA is transformed in culture into a high-affinity ligand for RXR α , then the overexpressed receptors in the nuclei would bind this ligand and trap it there.

We confirmed *in vitro* that RXR α does not bind [³H]*t*-RA. Binding studies in whole cells show that nucleosol fractions from [³H]*t*-RA-treated COS-1 cells transfected with either RAR α or RXR α bind similar amounts of radioactivity. To determine the nature of the ligand bound in RAR α - and RXR α -transfected COS-1 cells, we extracted labelled nucleosols with organic solvents¹⁶. High-performance liquid chromatography (HPLC) analysis of these organic extracts provided the first evidence that RXR α trapped a distinct derivative of *t*-RA. Extracts from RAR α -, RAR β - and RAR γ -transfected cells contain *t*-RA (elution time, 29.8 min, compare Fig. 1a and 1d peak e; data not shown). Extracts from RXR α -transfected nucleosols contain very little *t*-RA, but contain a tritiated material that elutes earlier than *t*-RA (Fig. 1b, elution time, 29.0 min). The RAR α and RXR α ligands are distinct chemical entities, because baseline resolution of the two substances is achieved after co-injection of the samples (Fig. 1c). After testing a number of standard compounds, we found that the RXR ligand elutes with the same retention time as 9-*cis* retinoic acid (9-*cis* RA) (compare Fig. 1b and d peak d). In addition, diazomethane derivatized¹⁷ ligands from RAR α and RXR α coelute with standards for all-*trans* methyl retinoate (Me-*t*-RA, elution time, 39.5 min), and 9-*cis* methyl retinoate (Me-9-*cis*-RA elution time, 38.6 min), respectively.

We used microbore liquid chromatography/mass spectrometry (LC/MS) with negative and positive chemical ionization to further characterize the RXR ligand¹⁸. Methylated derivatives of the RAR and RXR ligands were analysed using negative chemical ionization mass spectrometry and monitoring the total ion current (TIC) including ions of mass to charge ratio (*m/z*) 150–500. The TIC of the methylated derivatives of both the RAR and RXR ligands have only a molecular ion at *m/z* 314, and cochromatograph with Me-*t*-RA (data not shown) and

FIG. 1 Reversed-phase HPLC analysis of extracts of nucleosols from RAR α - and RXR α -transfected COS-1 cells treated with [³H]*t*-RA. a, The predominant peak of radioactivity in the organic extracts of the nucleosol fractions from RAR α -transfected cells elutes with a retention time of 29.8 min and coelutes with authentic *t*-RA standard (d, peak e). b, The predominant peak of radioactivity in the organic extract of nucleosol fractions from RXR α -transfected cells elutes with a retention time of 29.0 min. This peak coelutes with authentic standards for 9-*cis* RA (d, peak d). c, Coinjection of the organic extracts of nucleosol fractions from [³H]*t*-RA-treated COS-1 cells transfected with either RXR α or RAR α yields two distinct peaks of radioactivity. d, Retinoic acid isomer standards monitored at 351 nm have the following retention times (in min) peak a, 3,4 dihydro-RA (26.2); peak b, 13-*cis* RA (27.7); peak c, 11-*cis* RA (28.3); peak d, 9-*cis* RA (29.0); and peak e, all-*trans* RA (29.8). Other metabolites of retinoic acid elute at the following retention times, 4-hydroxy RA (12.8), and 4-oxo RA (14.4).

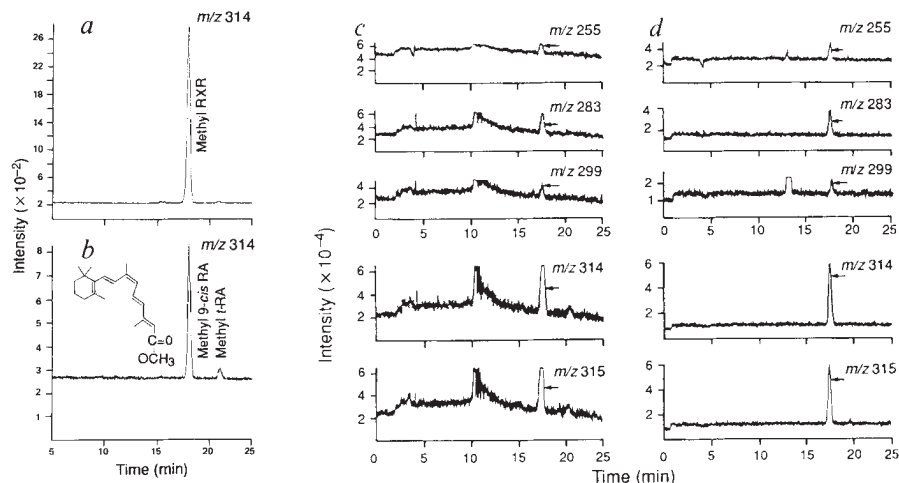
METHODS. COS-1 cells were cultured¹² and near-confluent cells were transfected with the pSG5 expression plasmid²⁹ containing complementary DNAs for RAR α , RXR α , or in the case of mock-transfected cells, a chimaeric human oestrogen receptor containing the DNA-binding region of RAR α (construct A, ref. 30). The human RAR α and RXR α coding sequences were cloned by polymerase chain reaction (PCR) from HL60 cell messenger RNA and human liver mRNA (Clontech), respectively, and inserted into pSG5. COS-1 cells (5×10^7) were washed, suspended in 0.8 ml of Ca²⁺- and Mg²⁺-free PBS, mixed with 25 μ g plasmid DNA, and electroporated (Bio-Rad GenePulser) at 250 μ F and 350 V. Transfected COS-1 cells were treated after 24 h with 50–200 nM [³H]*t*-RA and the cells were collected 48 h after drug treatment. The all-*trans*-[11, 12-³H]₂retinoic acid was obtained from NEN and checked for radiochemical purity (99.5%) by HPLC analysis. Nucleosol fractions were prepared¹² 72 h after transfection and were extracted with acetonitrile:butanol (1:1)¹⁶. The organic extracts were injected with a WISP model 712 autinjector in volumes of 45–100 μ l and eluted at a flow rate of 1.5 ml min⁻¹ from a series of two Zorbax ODS 4.6 mm $\times$ 25 cm columns using a linear gradient of acetonitrile: 20 mM ammonium acetate: acetic acid 50:50:0.5 to 95:5:0.04 in 26 min and holding final conditions for 12 min for free acids or 20 min for methyl esters. A flow-through radiochemical detector (IN/US β -RAM) with a 540 μ l flow cell and a scintillant flow rate of 4.5 ml min⁻¹ were used to determine the radioactivity in the eluent. Standard



curves for converting net counts on the radiochemical detector to d.p.m. were generated using HPLC analysis of known quantities of [³H]*t*-RA. All procedures using retinoids or retinoid-treated cells or cell fractions were done in darkness or amber light. Retinoic acid metabolite and isomer standards were synthesized in the Department of Chemistry, Hoffmann-LaRoche.

FIG. 2 Normal phase microbore LC/MS analysis of diazomethane derivatives of nucleosol extracts and standards. *a, b*, The negative chemical ionization reconstructed selected ion current profiles of diazomethane derivatives of RXR α nucleosol extract (*a*) has the same retention time (18.5 min) and molecular ion as authentic methyl-9-*cis* RA (*b*). *c, d*, Qualitative identification of the RXR α extract as 9-*cis* RA by positive chemical ionization. The ions monitored (arrows) were: m/z 315, 314, 299, 283 and 255 which were assigned $(M+H)^+$, $(M)^+$, $(M-CH_3)^+$, $(M-OCH_3)^+$, $(M-COOCH_3)^+$, respectively¹⁹. *c*, The reconstructed selected ion current profiles of the methyl ester of the RXR α extract. *d*, The reconstructed selected ion current profiles of the methyl ester of authentic 9-*cis* RA.

METHODS. Transfected COS-1 cells were treated with 500 nM *t*-RA 24 h after transfection with RAR α , mock, or RXR α plasmids and collected 48 h after treatment. Organic extracts of the nucleosol fractions were derivatized with diazomethane¹⁷, and fractionated by reversed-phase HPLC (see legend to Fig. 1) to collect the peaks of interest. The fractions corresponding to the RXR or RAR ligands were collected, evaporated to dryness and resuspended in mobile phase. Aliquots (9 μ l) were applied to a diol column (1 $\times$ 250 mm) (ES Industries) and eluted isocratically using a mobile phase of 15% toluene in *n*-hexane at a flow rate of 50 μ l min⁻¹ at 45 °C. The



eluent was introduced into a Finnigan 3200 mass spectrometer modified for direct liquid introduction³¹ through a probe heated to 250 °C. Mobile phase was used as the reagent gas for both positive and negative chemical ionization.

Me-9-*cis* RA, respectively (Fig. 2 *a, b*). We monitored ion fragments using positive chemical ionization MS with selected ion monitoring of the five fragment ions characteristic of methylated retinoic acid isomers¹⁹. The methylated RXR ligand and the Me-9-*cis* RA cochromatograph and produce the identical fragment ions with similar intensities (Fig. 2 *c, d*).

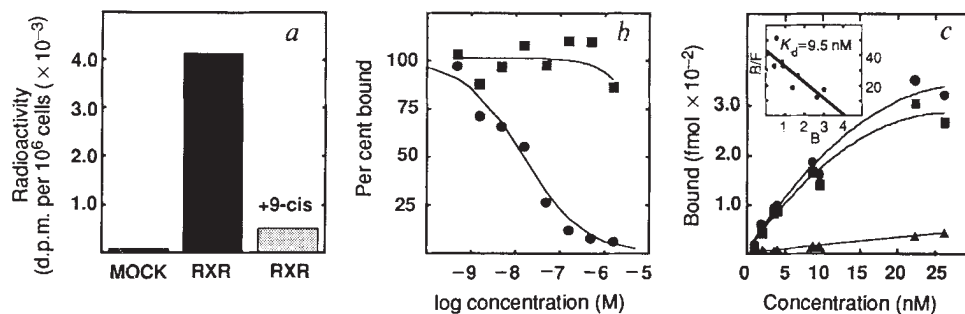
The RXR ligand and its methyl derivative coelute by both normal and reversed-phase HPLC with 9-*cis* RA and methyl 9-*cis* RA, respectively, and the fragmentation in both negative and positive chemical ionization between authentic Me-9-*cis* RA and the methylated derivative of the RXR ligand are identical. We propose that 9-*cis* RA is the ligand trapped by RXR α over-

expressed in COS-1 cells. We refer to the extracted RXR ligand as 9-*cis* RA(X). A chemical structure of the methyl derivative of the RXR ligand, 9-*cis* methyl retinoate is shown in Fig. 2 *b*.

In whole cell assays, [³H]9-*cis* RA(X) binding to macromolecules in nucleosol fractions from COS-1 cells is dependent on transfection with RXR α and is, effectively inhibited by cotreatment of the COS-1 cells with a 200-fold molar excess of unlabelled 9-*cis* RA (Fig. 3 *a*). Direct binding of 9-*cis* RA(X) in the nucleosol fraction of RXR α -transfected COS-1 cells was clearly demonstrated because HPLC analysis of organic extracts of these nucleosol preparations show trapping of [³H]9-*cis* RA (data not shown). *In vitro* binding studies demonstrate that

FIG. 3 *a*, Binding of the [³H]9-*cis* RA(X) in COS-1 cells transfected with RXR α or mock plasmid. [³H]9-*cis* RA(X) was isolated and used to treat COS-1 cells transfected with either mock plasmid or RXR α in the absence or presence of a 200-fold molar excess of unlabelled 9-*cis* RA. *b*, Competition for [³H]9-*cis* RA(X) binding *in vitro* to RXR α -enriched nucleosol by 9-*cis* RA (●) and *t*-RA (■). *c*, Saturation kinetics and Scatchard analysis of the binding of [³H]9-*cis* RA(X) to RXR α . Specific binding (■) is defined as the total binding (●) minus the nonspecific binding (▲). Scatchard analysis (inset) of the saturation kinetics represents bound (fmol)/free (nM) $\times 10^6$ versus bound (fmol).

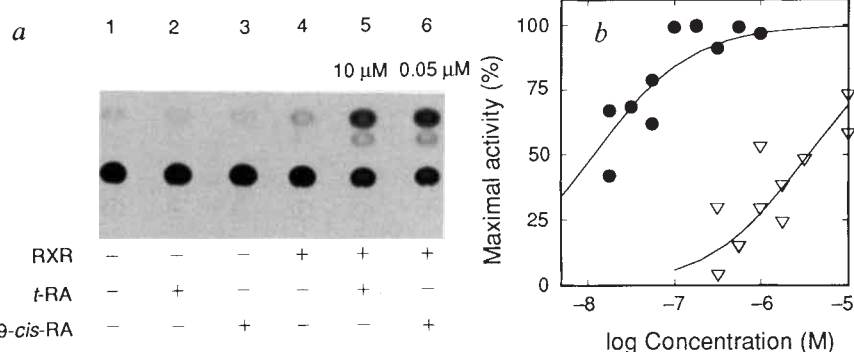
METHODS. COS-1 cells transfected with RXR α or mock plasmids were treated with [³H]9-*cis* RA(X) 48 h after transfection. The radioligand ([³H]9-*cis* RA(X)) was prepared by extracting the nucleosol fractions from RXR α -transfected cells treated for 48 h with 200 nM [³H]*t*-RA. The acetonitrile-butanol extract was concentrated by evaporation under a stream of nitrogen, reconstituted in ethanol and added to COS-1 cells to yield a final concentration in the medium of 0.4 nM. After 24 h, cells were collected and the nucleosol fractions were analysed to determine the bound activity. Bound was separated from free radioactivity by eluting the nucleosol from Pharmacia PD10 desalting columns with 5 mM phosphate buffer pH 7.4, 400 mM KCl, 10% glycerol, 1 mM dithiothreitol, and 1 mM phenylmethylsulphonyl fluoride. Bound radioactivity eluted in the column void volume and was determined by liquid scintillation counting of the eluent. The identity of the radiolabel in the nucleosol fractions of the [³H]9-*cis* RA(X)-treated cells



was determined following organic extraction and HPLC analysis as described in the legend to Fig. 1. *In vitro* binding studies with [³H]9-*cis* RA(X) were done incubating aliquots of RXR α nucleosol for 6 h at 4 °C in the nuclei lysis buffer¹² in a total volume of 50–100 μ l for competitive binding or saturation kinetics studies. The [³H]9-*cis* RA(X) was concentrated by evaporation under a stream of nitrogen to form the appropriate stock solutions. In competition binding studies, aliquots of RXR α nucleosol were incubated with 10 nM [³H]9-*cis* RA(X) in the presence of various concentrations of unlabelled *t*-RA or 9-*cis* RA. Saturation kinetics were done by adding increasing concentrations (1–26 nM) of [³H]9-*cis* RA(X) to RXR α nucleosol in the presence (nonspecific binding) or absence (total binding) of a 100-fold molar excess of unlabelled 9-*cis* RA. Bound radioligand was separated from free as described above. Calculations for the saturation kinetics and Scatchard analysis^{32,33} were based on the specific activity of the [³H]*t*-RA (50.6 Ci mmol⁻¹) used in the preparation of the radiolabelled extracts. Specific activity was confirmed by LC/MS analysis showing the presence of two tritium atoms in the RXR radioligand.

FIG. 4 Comparison of 9-*cis* RA and *t*-RA induction of RXR α -dependent transactivation of CRBP-II-RXRE-CAT in CV-1 cells. **a**, CV-1 cells were cotransfected with the reporter CRBP-II-RXRE-CAT construct (lanes 1–6) and either plasmid RXR α (lanes 4–6) or control plasmid Bluescript II KS (lanes 1–3). Cells were exposed for 48 h to either 50 nM 9-*cis* RA (lanes 3 and 6) or 10 μ M *t*-RA (lanes 2 and 5) or solvent (lanes 1 and 4). **b**, Concentration effect curves for the RXR α -dependent transactivation of CRBP-II-RXRE-CAT by 9-*cis* RA (●) and *t*-RA (▽) after 12 h of treatment.

METHODS. To construct CRBP-II-RXRE-CAT, oligos 5'-GATCTGCTGCACAGGTACACAGGTACACAGGTACAGTTCA-3' and 5'-GATCTGAACCTGTGACCTGTGACCTGTGACCTGTGACAGCA-3' (ref. 20) were annealed and inserted into the *Bgl*II site of the vector pCAT-Promoter (Promega) upstream of the SV40 minimal promoter to generate the reporter plasmid CRBP-II-RXRE-CAT. For transfection experiments, the RXR α expression plasmid (0.5 μ g) (except lanes 1–3 **a**), reporter plasmid CRBP-II-RXRE-CAT (5 μ g), reference plasmid p β GalactZ (5 μ g), and carrier plasmid Bluescript II KS (Stratagene) were combined for a total of 15 μ g and introduced into CV-1 cells (5×10^5 per 100-mm plate) by calcium phosphate coprecipitation³⁴. After 18 h, the precipitate was removed and cells washed in phosphate buffered saline. Various concentrations of 9-*cis* RA or *t*-RA were added to the medium after the initiation of



transfection. Cells were collected 48 h (**a**) or 12 h (**b**) after treatment, and lysates prepared and analysed for CAT activity³⁵. β -galactosidase activity was determined using Lumigal (Lumigen) according to the manufacturer's specifications. For autoradiographic presentation of the CAT activity, cell lysate volumes added to the assay tubes were normalized to β -galactosidase activity. For analysing concentration response relationships, CAT activity was quantitated directly from TLC plates using a Betascope 603 beta analyser (Betagen) and counts were normalized to β -galactosidase activity.

[³H]9-*cis* RA(X) exhibits saturable, high-affinity binding (dissociation constant, $K_d = 9.5$ nM) to RXR α in nucleosol prepared from RXR α -transfected COS-1 cells (Fig. 3c). Unlabelled 9-*cis* RA (50% inhibitory concentration, IC_{50} 20–50 nM), but not *t*-RA ($IC_{50} > 10$ μ M) competes for [³H]9-*cis*-RA(X) binding to RXR α nucleosol preparations (Fig. 3b). Thus, 9-*cis*-RA binds directly to RXR α in whole cells and *in vitro* has a relatively high affinity for this receptor.

To test whether 9-*cis* RA can activate RXR α , we constructed a reporter plasmid (CRBP-II-RXRE-CAT) with the CRBP-II RXR response element²⁰. Cotransfection of CV-1 cells with RXR α and the CRBP-II-RXRE-CAT reporter gene results in an increase in chloramphenicol acetyltransferase (CAT) activity that is dependent on the addition of either *t*-RA or 9-*cis* RA (Fig. 4a). Similar CAT activity is obtained when CV-1 cells are treated with 0.05 μ M 9-*cis* RA or 10.0 μ M *t*-RA for 48 h before assay (Fig. 4a). We attribute the potency of *t*-RA at this time point to the isomerization of *t*-RA to 9-*cis* RA in the CV-1 cells. To minimize isomerization, CV-1 cells were treated with 9-*cis* RA or *t*-RA for only 12 h before assaying for CAT activity. Under these conditions, there is a marked difference in the potencies of 9-*cis* RA and *t*-RA in inducing CAT activity, with these compounds having EC_{50} s (effective concentrations for half maximal response) of 18 nM and > 1 μ M, respectively (Fig. 4b).

We demonstrate that the 9-*cis* stereoisomer of retinoic acid directly binds and activates RXR α and we propose that 9-*cis* RA, is the cognate ligand for the RXR subfamily of retinoid receptors. Because RXR α does not bind *t*-RA, isomerization of *t*-RA to 9-*cis* RA could control RXR-specific gene activation and represent a new pathway through which retinoic acid can control cell physiology. Although 9-*cis* RA alone directly activates the RXR subfamily, it seems that both 9-*cis* RA and *t*-RA can directly activate the RAR subfamily with similar potency (M. Saunders, M.-T. Bocquel and P. Chambon personal communication). Clearly, a simple interaction of one isomer with each subfamily cannot explain the activities mediated by retinoid receptors. But it is becoming clear that complex interactions among the receptors themselves, or with other members of this superfamily^{21–25}, and with other nuclear protein factors^{7,26,27} are responsible for controlling the overall molecular mechanism of retinoid action. Recent evidence has shown that RXRs form heterodimers with RARs and this interaction enhances transcriptional activation on RAR response elements²⁸. Conceivably, transcriptional activation of retinoid receptors could depend in

part on an isomer exchange in an RXR-RAR heterodimer complex. Retinoid action in general is likely to be controlled by the presence of a variety of receptor subtypes, the formation of receptor heterodimers, interaction with other transcription factors, and as we suggest here, the specific isomers of retinoic acid available in the ligand pool. □

Received 27 November 1991; accepted 8 January 1992.

- Thaller, C. & Eichele, G. *Nature* **327**, 625–628 (1987).
- Roberts, A. B. & Sporn, M. B. *The Retinoids* Vol 2 (eds Sporn, M. B., Roberts, A. B. & Goodman, D. W. S.) 209–286 (Academic, Orlando, 1984).
- Coward, W. A., Howell, J. M., Thompson, J. N. & Pitt, G. A. *J. Br. J. Nutr.* **23**, 619–626 (1969).
- Giguere, V., Ong, E. S., Segui, P. & Evans, R. M. *Nature* **330**, 624–629 (1987).
- Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. *Nature* **330**, 444–450 (1987).
- Åström, A., Pettersson, U., Krust, A., Chambon, P. & Vorhees, J. J. *Biochem. biophys. Res. Comm.* **173**, 339–345 (1990).
- Yang, N., Schüle, R., Mangelsdorf, D. J. & Evans, R. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3559–3563 (1991).
- Graupner, G. *et al. Biochem. biophys. Res. Comm.* **179**, 1554–1561 (1991).
- Brand, N. *et al. Nature* **332**, 850–853 (1988).
- Krust, A., Kastner, Ph., Petkovich, M., Zentgraf, A. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5310–5314 (1989).
- Mangelsdorf, D. J., Ong, E. S., Dyck, J. A. & Evans, R. M. *Nature* **345**, 224–229 (1990).
- Nervi, C. J., Grippo, J. F., Sherman, M. I., George, M. D. & Jetten, A. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5854–5858 (1989).
- Cavey, M. T., Martin, B., Caravan, I. & Shroot, B. *Analyt. Biochem.* **186**, 19–23 (1990).
- Crettaz, M., Baron, A., Siegenthaler, G. & Hunziker, W. *Biochem. J.* **272**, 391–397 (1990).
- Delescluse, C. *et al. Molec. Pharmacol.* **40**, 556–562 (1991).
- McClean, S. W., Ruddel, M. E., Gross, E. G., DeGiovanna, J. J. & Peck, G. L. *Clin. Chem.* **28**, 693–696 (1982).
- De Leenheer, A. P. & Lambert, W. E. *Meth. Enzym.* **189**, 104–111 (1990).
- Husellon, C. A. *et al. in Liquid Chromatography/Mass Spectrometry, Applications in Agricultural, Pharmaceutical and Environmental Chemistry* (ed. Brown, M. A.) 166–178 (American Chemical Society, Washington DC, 1990).
- Napoli, J. L. *Meth. Enzym.* **123**, 112–124 (1986).
- Mangelsdorf, D. J. *et al. Cell* **66**, 555–561 (1991).
- Glass, C. K., Lipkin, S. M., Devary, O. V. & Rosenfeld, M. G. *Cell* **59**, 697–708 (1989).
- Husmann, M. *et al. Molec. cell. Biol.* **11**, 4097–4103 (1991).
- Hudson, L. G., Santon, J. B., Glass, C. K. & Gill, G. N. *Cell* **62**, 1165–1175 (1990).
- Zhang, X. K. *et al. New Biol.* **2**, 169–181 (1991).
- Graupner, G., Willis, K. N., Tzukerman, M., Zhang, X.-K. & Pfahl, M. *Nature* **340**, 653–656 (1989).
- Schüle, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 6092–6096 (1991).
- Glass, C. K., Devary, O. V. & Rosenfeld, M. G. *Cell* **63**, 729–738 (1990).
- Yu, V. C. *et al. Cell* **67**, 1251–1266 (1991).
- Green, S., Isseman, I. & Scheer, E. *Nucleic Acids Res.* **16**, 369 (1988).
- Dalman, F. C. *et al. Biochemistry* **30**, 5605–5608 (1991).
- Garland, W. A. *et al. Trends analyt. Chem.* **3**, 177–184 (1991).
- Scatchard, G. *Ann. N.Y. Acad. Sci.* **51**, 660–672 (1949).
- Grippo, J. F. & Gudas, L. J. *J. biol. Chem.* **262**, 4492–4500 (1987).
- Chen, C. & Okayama, H. *Molec. cell. Biol.* **7**, 2745–2752 (1987).
- Rosenthal, N. *Meth. Enzym.* **152**, 704–720 (1987).

ACKNOWLEDGEMENTS. We thank F. Vane, D. Liberato and B. Wolitzky for advice on techniques, D. Lucas and S. Penrick for plasmid DNA, E. A. Pfutzer, R. A. Salvador and R. K. Kuntzman for their support and discussions, P. Chambon for the pSG5 expression vector and the chimaeric receptor used in mock transfections and G. Vasios for the plasmid p β GalactZ.

Herpesvirus saimiri encodes homologues of G protein-coupled receptors and cyclins

John Nicholas, Keith R. Cameron & Robert W. Honess

Division of Virology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

HERPESVIRUS saimiri (HVS) is a T-lymphotropic gammaherpesvirus which establishes asymptomatic infections in its natural host the squirrel monkey (*Saimiri sciureus*), but which causes fatal lymphoproliferative diseases in other New World primates¹. Sequencing studies show HVS is closely related to the human B-lymphotropic gammaherpesvirus Epstein-Barr virus (EBV)²⁻⁴. However, despite the general collinearity between the genomes of HVS and EBV, HVS contains genes not found in EBV or in the genomes of any of the other sequenced herpesviruses⁵⁻⁸. We have identified two genes, occurring in a region of divergence between HVS and EBV, that have cellular homologues. One of these, *ECRF3*, is homologous to the genes encoding the human cytomegalovirus (HCMV) and cellular G protein-coupled receptor family of proteins⁹. The other HVS gene, *ECLF2*, is homologous to the genes encoding cellular cyclins and to our knowledge is the first reported example of a viral cyclin. The presence of G protein-coupled receptor and cyclin homologues in HVS suggests that these genes may be important in the regulation of viral and cellular processes during productive and/or latent infection of host cells, and in particular may be of relevance in the transformation and rapid proliferation of T cells during HVS infections of hosts susceptible to HVS-induced lymphoproliferative diseases.

The sequence presented in Fig. 1 (approximate map units 0.93-0.96) represents part of a larger contiguous sequence comprising 43.6 kilobase pairs (kb) from the right end of the unique component (L-DNA) of the HVS genome (J.N., R.W.H., H. Coleman and K.R.C., manuscript in preparation). The *ECLF2* and *ECRF3* translation products (starting at the first methionine) are shown above the nucleotide sequence, and contain 254 and 321 amino acids, respectively, with predicted relative molecular masses of 28,600 and 37,100 (28.6K and 37.1K). Although we have no direct evidence for the expression of either of these two genes, pseudogenes have not been identified in other herpesviruses, the first ATG codons conform to the Kozak consensus¹⁰ and there is a polyadenylation signal (AATAAA) immediately downstream of the *ECRF3* reading frame (position 39,414).

The *ECRF3* translation product was used to screen the NBRF database for potential homologues using a FASTA search¹¹. The results of the search showed only one significant match (score of 225; 22% identity over a 295 amino-acid overlap) with the HCMV-US28 protein, homologous to the opsin family of cellular receptors, together with the products of the HCMV US27 and UL33 genes⁹. Pairwise alignments between the HVS *ECRF3* protein and each of the HCMV G protein-coupled receptor (GCR)-homologues and selected cellular members of the GCR family allowed the homology of *ECRF3* protein with each of these proteins to be detected and compared (Fig. 2 and Table 1a). The percentage similarities and identities between *ECRF3* protein and the HCMV and cellular proteins are comparable and represent values obtained from complete protein overlaps,

TABLE 1 Amino-acid sequence conservation

(a)	US27	US28	UL33	5HTA	mas	BRh
ECRF3	46.5 (17.2)	44.0 (21.0)	47.7 (19.4)	46.0 (16.2)	48.0 (17.7)	45.9 (20.2)
HCMV-US27		58.9 (31.5)	46.2 (21.5)	42.1 (19.7)	48.9 (21.4)	40.9 (15.6)
HCMV-US28			47.8 (19.8)	51.3 (27.3)	49.3 (18.5)	49.8 (20.1)
HCMV-UL33				37.0 (19.2)	48.2 (18.1)	44.8 (19.0)
5HTA					46.7 (21.3)	47.9 (22.8)
mas						47.6 (18.4)

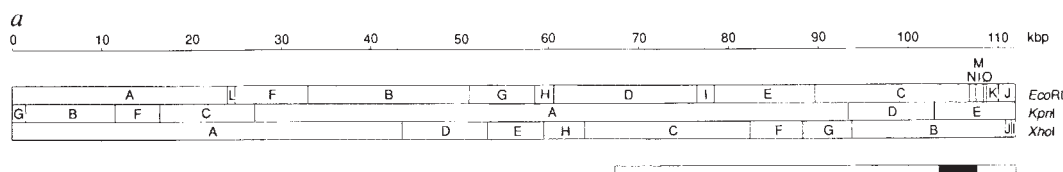
(b)	Clam A	DmA	DmB	XIA	HuD
ECLF2	40.6 (20.1)	42.1 (22.0)	43.2 (24.0)	37.0 (19.3)	45.7 (23.2)
ClamA		52.3 (34.0)	45.1 (25.7)	60.6 (43.3)	45.9 (23.3)
DmA			45.1 (25.7)	51.9 (34.5)	47.1 (23.5)
DmB				43.4 (25.4)	44.8 (23.6)
XIA					42.6 (21.6)

a, Comparisons of conservation between HVS-*ECRF3* translation product, HCMV-US27, HCMV-US28, HCMV-UL33 (ref. 9), 5-hydroxytryptamine 1A receptor (5HTA)³¹, human *mas* oncoprotein²⁵ and bovine rhodopsin (BRh)³² amino-acid sequences. Values for per cent similarity and per cent identity (in parentheses) were derived from the output of GAP³³ pairwise alignments with gap and length weights set at 3.0 and 0.1, respectively. b, Comparisons of amino-acid sequence conservation between the HVS-*ECLF2* translation product, clam cyclin A (clamA)²¹, *Drosophila* cyclin A (DmA)³⁴, *Drosophila* cyclin B (DmB)³⁵, *Xenopus* cyclin A (XIA)²⁰ and human cyclin D1 (HuD)¹⁹. Gap and length weights were set at 5.0 and 0.3, respectively.

not from overlaps merely of the best conserved regions. The cellular and HCMV GCR-type proteins are characterized by seven hydrophobic regions that are believed to be transmembrane domains^{12,13}, and similar hydrophobic regions are also detected in the *ECRF3* protein (Fig. 2).

In addition to the sequence and hydrophobicity similarities, the *ECRF3* protein shares other features characteristic of GCRs. These include glycosylation sites in the N-terminal putative extracellular domain that are found in most cellular GCRs and two (US27 and US28) of the HCMV GCR homologues (Fig. 2), where they may have a role in receptor localization, although they seem not to be required for receptor function¹⁴. Also conserved are proline residues in putative transmembrane regions II, IV, V, VI and VII, and two cysteine residues, in the putative second and third extracellular domains, which are found in all the characterized GCRs and which in bovine rhodopsin are the only two of 10 cysteines that are structurally essential¹⁵.

Approximately 2 kb upstream of *ECRF3* is the *ECLF2* reading frame encoding a product homologous to the cyclins characterized from a variety of organisms¹⁶⁻¹⁹. Cyclins are found in association with cdc2 or related protein kinases, and their activity and function in mediating cell-cycle progression are absolutely dependent on this interaction^{16,17,20-22}. The cyclins



C O H L C I D W T S R S L T E I I D S I N V A N C S R P I C G N I L F L A
 T A A C A C T G C A T T A G C A A A T C C C A T G T G A T T T T T G A T A A G T T T A T A T T A A A T C G A G A T T A A C T G C A T T A C A G C T T T G U G A A T A C C A C C A T T A T A A A A A T A A T A G
 35709 35719 35729 35739 35749 35759 35769 35779 35789 35799 35809 35819
 C Y M H E E D T F S A D T I H L V T T R L D M L I E L D Y L S F A A E S Q D K V T
 A C A G T A C C T T T C A A T T G T T A A A T G C T G T T A T G T T A C A G C A G T T T T T A A G C T A T T T A C A G A A T T T G A G C A G A A G T T C A G A C A G C T T T T T T A C T T C T
 35829 35839 35849 35859 35869 35879 35889 35899 35909 35919 35929 35939
 R V T N T S F N L I S S L D E L Y C T M P R C H T H D T E T I T T L L G G A C I L
 T C T T A C T G T T A G A G A C T T A A A T A A G A A A A G G C T T T A A G T A A C A G C T G A A G C T G C T A G C T T T T C A G T T T C A A A T G T T T T A A C A G T C T C T G C A C A A A T A G
 35949 35959 35969 35979 35989 35999 36009 36019 36029 36039 36049 36059
 G P S L L A I N P O I L A R C I T T S A E Y F L O P M L D E P I R L A N C L P I
 A C C T G G A C A C A A G A G A T T C T G G G T T A A T A A A G C T T G C A A A T A G A T A G T A G C T T A C A G A G A G G C A A A G C T T T T G G T A T T T A A A G C A T A T A C A C G T G C A A T
 36069 36079 36089 36099 36109 36119 36129 36139 36149 36159 36169 36179
 L F D T A L V A E T D M K L A E L I O K E N I G L E L N T F C D C S L Y T L R S
 T A A A A C T T G T A G A A G T A C A G C T A G T G C T C C A T T C A G C T T T A A A A G T C T T T T T C T G G T A G A T C T A A A T T T T A A A C A G A C A A G A A A A G T C A G T T T T T T C A
 36189 36199 36209 36219 36229 36239 36249 36259 36269 36279 36289 36299
 V T M P K V S G I L V C A A G I K Q L T K R T G Q K X C L Y R D L I S
 C A C A G T A T T G G T T T A C A G T T C T A T C T T C T A C A A G A C A G C T C C T A T T T T T C T A A G T C T T T T T T A G T G C T T G C T T T T T A C A G A T A T C T T G T T T A T T C A T A G A C T T A G
 36309 36319 36329 36339 36349 36359 36369 36379 36389 36399 36409 36419
 S L P F V S K D L E F S E C L L H M W T L L I T R N D V T V E T Q I E W L S T F
 A C T G A G G A A A C A C T T T C T A A C T A A A A G T T T C A C A G A C A G T C A A G T A A G T T T A T T G T C A A A G A T T T C T T T T T A T T C A T A G A C T A G C T A G T A A A
 36429 36439 36449 36459 36469 36479 36489 36499 36509 36519 36529 36539
 K P L L L E R L K L L N H L V R P D K M T T S D I K A R N L R H P S D G C
 T T T T G G A A G A A C G T T C G A G T T T T A A G T T T T A G A C T T T G T G C T T T A C T A G T A G A C T T A T T T T A G C T T G T G A G T T T T T T G C T A A T
 36549 36559 36569 36579 36589 36599 36609 36619 36629 36639 36649 36659
 K I P L S F S F R G L S A K L K A S R P Y K R E Y
 C T G A A A T C A G C T T A G C T A G T T T T A A A T G C T T C A T G A T C A T T T T A T A G G A A G C T T T T T G C T A A T T T T C A A A G A G A T C T T T A A T T T T A G A C A T A T A G G A T T T T T T T T C T G
 36669 36679 36689 36699 36709 36719 36729 36739 36749 36759 36769 36779
 K K Q A K I V E A K A T O F T R C Y V T I F Y V M P G T A N P G X P F T Q I G S Y
 A T T T T T T C A G C T T C A T A C T T T T T A G C A T T T T G A A T G T T C T C A C A C A G T A T A A A A T A C C A A G A T T A C C A A G A G T C T T T A A G A A T T T T T T T A T T T T T T T C C A A A G
 36789 36799 36809 36819 36829 36839 36849 36859 36869 36879 36889 36899
 V A G F F S L S R S L R Y L A D T P G M F D M W C S K P S L D L V S P A A O
 A A A C C T G G G A A C A A A C T T A A C T T G S A G T A C A G A C A C C A G T T G G C C A A A A A T T C A C A C A A C A A G A C A A G T T T T T T A A A C A A G A C A T A G C T T A G C T G C C T T
 36909 36919 36929 36939 36949 36959 36969 36979 36989 36999 37009 37019
 S F F K N R A Q Y K K P V R D H S Q R G R P L R S R H Q P Q M R C C L E P V G G
 G A C T A A C C A T T A T T T T T T T A T T T T A G G A A C T T T A C C T C A C T G T T T T C T T T T T T G C C A G A C A G G T G C T A G A G T T T T T T T A C A C A A T T T T T G T C A A C A A T T T T T G T C A
 37029 37039 37049 37059 37069 37079 37089 37099 37109 37119 37129 37139
 P O T S P R O G V V R Y H P L R P T S P C A E Z E E E E E A E E A E E A E E A E
 C T G C T T A G T A G A C G T T T T T T C A T T T T G A T T T G T T A A G C T G A C T G C T G C T T T C T T C T T C T A G C T T T C A G C T T T T C A G C T T T T C T T C T T C T T C T T A G C T T
 37149 37159 37169 37179 37189 37199 37209 37219 37229 37239 37249 37259
 E A E A E A E E E A E E A E E A E E A E E A E E A E E A E E A E E A E E A E
 C T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T
 37269 37279 37289 37299 37309 37319 37329 37339 37349 37359 37369 37379
 E A E E E A E E A E E A E E E A E E A E E A E E A E E A E E A E E A E E A E
 C T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T C A G C T T T
 37389 37399 37409 37419 37429 37439 37449 37459 37469 37479 37489 37499
 E E E E E A E E A E E A E E A E E A E E A E E A E E A E E A E E A E E A R K E
 C T T C T C T T C A G C T T T C T T C T T C A G C T T T C T T C A G C T T T C T T C T T C A G C T T T C T T C T T C A G C T T T C T T C T T C A G C T T T C T T C T T C A G C T T T C T T C T T C A G C T T T C T T
 37509 37519 37529 37539 37549 37559 37569 37579 37589 37599 37609 37619
 E A E E E A E E R G E G E E G G E G E R E E E G R R E R E E E G E E E E E R
 C T T C A G C T T T C T T C T T C A G C T T T C T T
 37629 37639 37649 37659 37669 37679 37689 37699 37709 37719 37729 37739
 Q E T E L A A Q Q Q H T T N I D D G Q P K L K R R R K R P S V Y C O V H C K C
 C T T T C T T C T T T A G C T T C T T C T T T G C C G C T T T T A A T G T C T G C C G C T T G A G C T T A G T T T T C T T C T T C T T C T T C T T C T T C T T C T T C T T C T T C T T C T T C T T C T T C T T C T T
 37749 37759 37769 37779 37789 37799 37809 37819 37829 37839 37849 37859
 K D K C E S R L T H R R R K A K R R R P A H
 A T T T G C T T T C A G C A G A A C A G A G T T G C C T T T T T T T

ECLF2

ECRF3

FIG. 1. *a*, Diagrammatic representation of the L-DNA (unique component) region of the HVS genome showing *EcoRI*, *KpnI* and *XhoI* cleavage sites. The location of the 43.6-kb contiguous sequence from which the data presented here were derived (J.N., R.W.H., H. Coleman & K.R.C., manuscript in preparation) is indicated by the open box, and the 3.8-kb sequence (positions 35,700–39,500) containing the *ECLF2* and *ECRF3* genes is indicated by the filled box. *b*, Nucleotide sequence of part of the HVS genome containing the *ECLF2* and *ECRF3* genes, and predicted protein sequences (single-letter amino-acid code) of these and the intervening *ECLF1* gene. The positions and orientations of the *ECLF2* and *ECRF3* genes are indicated by arrows, and the polyadenylation signal (AATAAA) just downstream of the *ECRF3* gene, at position 39,414, is underlined. Sequences of both strands were determined by M13 shotgun DNA sequencing using the dideoxy chain termination method³⁶, and assembled and analysed as described previously².

accumulate during successive cell divisions as the result of rapid proteolytic degradation of the proteins at specific times before mitotic division, and different cyclin subtypes have distinct kinetics of accumulation^{16-18,20}.

The characterized cyclins, types A, B, C and D, are fairly well conserved both between themselves and between species (see for example, Table 1b), although different cyclins show greater divergence in amino-acid sequence towards their N termini, and the lengths of these less-conserved regions vary somewhat. Notably, the HVS *ECLF2* translation product lacks this N-terminal region, and thus is comparable in length to the D-type cyclins, and displays much less similarity with the cellular cyclins over analogous regions of these proteins than exists between the cellular cyclins. However, as shown in the alignments between the *ECLF2* translation products and human cyclin D, *Drosophila* cyclin A, *Drosophila* cyclin B, and *Xenopus* cyclin A (Fig. 3), the *ECLF2* gene is clearly homologous to the cyclins. The similarities between these proteins (and clam cyclin A) can be seen from the values shown in Table 1b.

The presence of homologues of the GCR and cyclin families

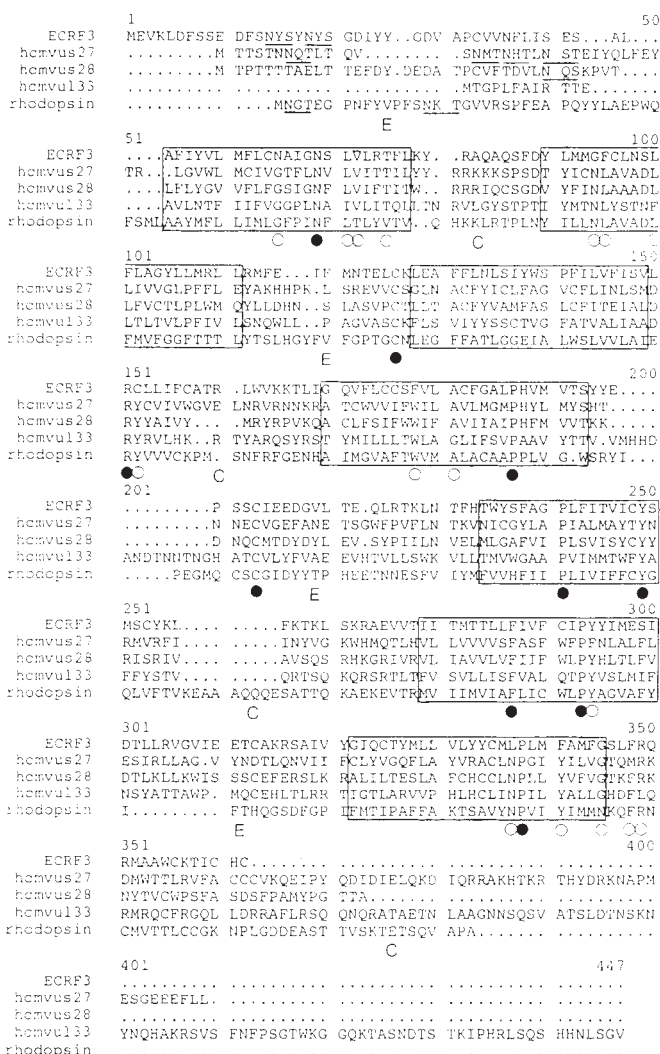


FIG. 2 Alignment of the *ECLF3* translation product with each of the three HCMV GCR homologues, US27, US28 and UL33 (ref. 9), and bovine rhodopsin (rhodopsin)³². The alignments were based on those performed by Chee *et al.*⁹ and on pairwise alignments using GAP³³, and were aligned manually by using the LINEUP and PRETTY programs in GCG¹¹. The boxed regions indicate putative transmembrane domains and expected extracellular (E) and cytoplasmic (C) domains are labelled. Potential N-linked glycosylation sites occurring in the N-terminal regions of the proteins are underlined. ●, Amino acids conserved in all five proteins; ○, those conserved in all but one of the sequences.

of cellular proteins in HVS, an oncogenic herpesvirus effecting rapid proliferation of latently infected T cells in susceptible New World primates¹, is potentially of great significance as both of these sets of proteins are known to be involved in the regulation of cellular proliferation, differentiation and gene expression. Although the transforming potential of HVS correlates with the presence of an intact open reading frame encoding the so-called saimiri transformation-associated protein (STP)⁷, this gene has not been shown to be sufficient for transformation and it is possible, therefore, that other viral genes may also be involved in transformation and/or accelerated T-cell proliferation.

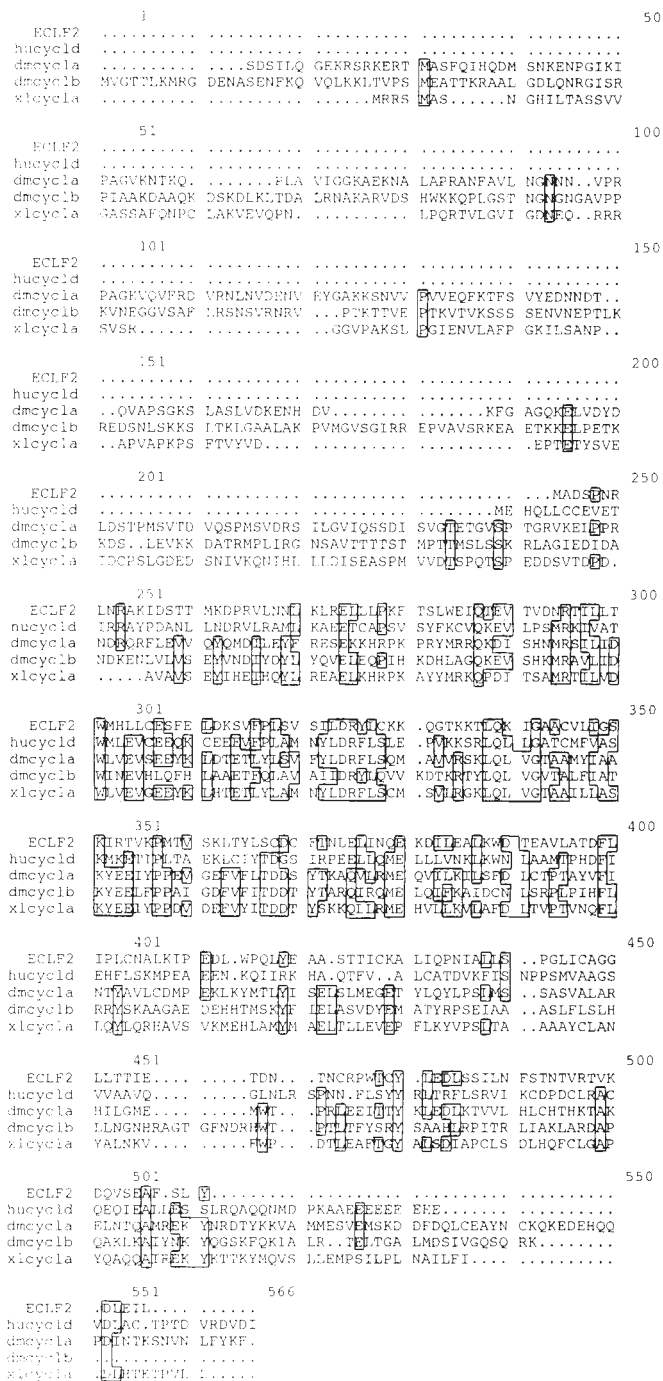


FIG. 3 Alignment of the *ECLF2* translation product with human cyclin D1 (hucyclD)¹⁹, *Drosophila* cyclins A (dmcyA)³⁴ and B (dmcyB)³⁵ and *Xenopus* cyclin A (xlcyclA)²⁰. The alignments of the five proteins are based on the outputs from pairwise alignments³³ and were assembled using the LINEUP and PRETTY options in GCG¹¹. Amino-acid residues conserved in at least three of the five proteins are boxed.

Specific members of the GCR protein family have a role in malignant transformation. For example, the ectopic expression of serotonin receptor (5HT1c) in NIH3T3 cells leads to serotonin-dependent transformation of these cells which can give rise to tumours in nude mice²³, and the human *mas* oncogene, which encodes an angiotensin receptor, is also tumorigenic in nude mice^{24,25}. Cyclins are required for cellular division, and thus have a key role in cellular proliferation and development¹⁶. Moreover, one of the cyclins, human D1 (ref. 19), is the *PRAD1* candidate oncogene implicated in the development of certain parathyroid tumours and hepatocarcinomas^{26,27}, and it is the D-type cyclins to which the HVS

ECLF2 protein seems to be most closely related. Cyclin A can associate with the cellular transcription factor E2F (refs 28, 29) which is likely to be directly involved in the regulation of expression of genes important for cellular proliferation and development, for example *c-myc*, *c-myb* and the EGF receptor gene³⁰. It is possible that the HVS cyclin may also associate with cellular transcription factors and thereby influence the expression of cellular and viral genes. Thus, the presence of cyclin and GCR homologues, *ECLF2* protein and *ECRF3* protein, in HVS may be highly relevant to the process of cellular transformation and rapid T-cell proliferation effected by HVS during latent infections of T cells in susceptible hosts. □

Received 16 September; accepted 22 October 1991.

1. Fleckenstein, B. & Desrosiers, R. C. in *The Herpesviruses* Vol. 1, 253-332 (ed. Roizman, B.) (Plenum, New York, 1982).
2. Cameron, K. R. et al. *J. Virol.* **61**, 2063-2070 (1987).
3. Gompels, U. A., Craxton, M. A. & Honess, R. W. *J. Virol.* **62**, 757-767 (1988).
4. Nicholas, J., Coles, L. S., Newman, C. & Honess, R. W. *J. Virol.* **65**, 2457-2466 (1991).
5. Honess, R. W. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3604-3608 (1986).
6. Trimble, J. J., Murthy, S. C. S., Bakker, A., Grassman, R. & Desrosiers, R. C. *Science* **239**, 1145-1147 (1988).
7. Murthy, S. C. S., Trimble, J. J. & Desrosiers, R. C. *J. Virol.* **63**, 3307-3314 (1989).
8. Nicholas, J., Smith, E. P., Coles, L. & Honess, R. *Virology* **179**, 189-200 (1990).
9. Chee, M. S., Satchwell, S. C., Preddie, E., Weston, K. M. & Barrell, B. G. *Nature* **344**, 774-777 (1990).
10. Kozak, M. *Nucleic Acids Res.* **9**, 5233-5252 (1981).
11. Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387-395 (1984).
12. Nathans, J. *A. Rev. Neurosci.* **10**, 163-194 (1987).
13. O'Dowd, B. F., Lefkowitz, R. J. & Caron, M. G. *A. Rev. Neurosci.* **12**, 67-83 (1989).
14. Flesler, S. J. & Basinger, S. F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1116-1120 (1985).
15. Karnik, S. S., Sakmar, T. P., Chen, H.-B. & Khorana, H. G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8459-8463 (1988).
16. Murray, A. W. & Kirschner, M. W. *Science* **246**, 614-621 (1989).
17. Wittenberg, C., Sugimoto, K. & Reed, S. I. *Cell* **62**, 225-237 (1990).
18. Matsushima, H., Roussel, M. F., Ashmun, R. A. & Sherr, C. J. *Cell* **65**, 701-713 (1991).

19. Xiong, Y., Connolly, T., Futcher, B. & Beach, D. *Cell* **65**, 691-699 (1991).
20. Minshull, J., Golsteyn, R., Hill, C. S. & Hunt, T. *EMBO J.* **9**, 2865-2875 (1990).
21. Swenson, K. L., Farrell, K. M. & Ruderman, J. V. *Cell* **47**, 861-870 (1986).
22. Draetta, G. et al. *Cell* **56**, 829-838 (1989).
23. Julius, D., Livelli, T. J., Jessell, T. M. & Axel, R. *Science* **244**, 1057-1062 (1989).
24. Young, D., Waitches, G., Birchmeier, C., Fasano, O. & Wigler, M. *Cell* **45**, 711-719 (1986).
25. Jackson, T. R., Blair, L. A. C., Marshall, J., Goerdert, M. & Hanley, M. R. *Nature* **335**, 437-440 (1988).
26. Wang, J., Chenivresse, X., Henglein, B. & Bréchet, C. *Nature* **343**, 555-557 (1990).
27. Motokura, T. et al. *Nature* **350**, 512-515 (1991).
28. Mudryj, M. et al. *Cell* **65**, 1243-1253 (1991).
29. Bandara, L. R., Adamczewski, J. P., Hunt, T. & La Thangue, N. B. *Nature* **352**, 249-251 (1991).
30. Mudryj, M., Hiebert, S. W. & Nevins, J. R. *EMBO J.* **9**, 2179-2184 (1990).
31. Fargn, A. et al. *Nature* **335**, 358-360 (1988).
32. Nathans, J. & Hogness, D. S. *Cell* **34**, 807-814 (1983).
33. Needleman, S. B. & Wunsch, C. D. *J. molec. Biol.* **48**, 444-453 (1970).
34. Lehner, C. F. & O'Farrell, P. H. *Cell* **56**, 957-968 (1989).
35. Lehner, C. F. & O'Farrell, P. H. *Cell* **61**, 535-547 (1990).
36. Barkier, A. T. & Barrell, B. G. in *Techniques in the Life Sciences* Vol. B5 (ed. Flavell, R. A.) 1-33 (Elsevier, Amsterdam, 1983).

ACKNOWLEDGEMENTS. We thank M. Craxton for discussions and critical reading of the manuscript and T. Hunt for advice.

S-phase feedback control in budding yeast independent of tyrosine phosphorylation of p34^{cdc28}

Peter K. Sorger* & Andrew W. Murray†

* Department of Microbiology and Immunology and † Department of Physiology, University of California, San Francisco, California 94143-0502, USA

IN somatic cells, entry into mitosis depends on the completion of DNA synthesis. This dependency is established by S-phase feedback controls that arrest cell division when damaged or unreplicated DNA is present¹. In the fission yeast *Schizosaccharomyces pombe*, mutations that interfere with the phosphorylation of tyrosine 15 (Y15) of p34^{cdc2}, the protein kinase subunit of maturation promoting factor, accelerate the entry into mitosis and abolish the ability of unreplicated DNA to arrest cells in G2 (ref. 2). Because the tyrosine phosphorylation of p34^{cdc2} is conserved in *S. pombe*³, *Xenopus*⁴, chicken⁵ and human⁶ cells, the regulation of p34^{cdc2}-Y15 phosphorylation could be a universal mechanism mediating the S-phase feedback control and regulating the initiation of mitosis^{7,8}. We have investigated these phenomena in the budding yeast *Saccharomyces cerevisiae*. We report here that the *CDC28* gene product (the *S. cerevisiae* homologue of *cdc2*) is phosphorylated on the equivalent tyrosine (Y19) during S phase but that mutations that prevent tyrosine phosphorylation do not lead to premature mitosis and do not abolish feedback controls. We have therefore demonstrated a mechanism that does not involve tyrosine phosphorylation of p34 by which cells arrest their division in response to the presence of unreplicated or damaged DNA. We speculate that this mechanism may not involve the inactivation of p34 catalytic activity.

We first established that p34^{CDC28} in budding yeast and p34^{cdc2} in fission yeast are phosphorylated on homologous

tyrosine residues. We prepared extracts from *S. cerevisiae* cells arrested in S phase with hydroxyurea, precipitated p34, and analysed it by immunoblotting. When precipitates from cells carrying wild-type (wt) *CDC28* were immunoblotted with anti-phosphotyrosine antibodies, a band corresponding to relative molecular mass of 34,000 (*M*_r 34K) was detected (Fig. 1b). In extracts from cells carrying a modified *CDC28* gene that encodes a fusion between p34^{CDC28} and a short epitope tag⁹ (Fig. 1a), the 34K band was replaced by a 38K band. To show the specificity of the immunoblotting and to localize the phosphorylated site, we constructed a mutant (*cdc28-F*) in which a phenylalanine was substituted for tyrosine 19 (Y19), the residue in p34^{CDC28} that is homologous to the phosphorylated tyrosine 15 of p34^{cdc2} in *S. pombe* (Fig. 1a)³. When p34 was analysed in extracts from *S. cerevisiae* cells carrying *cdc28-F* in place of wt *CDC28*, no phosphotyrosine was detected (Fig. 1b, lane 2) even though p34 was precipitated efficiently in all cases (lanes 5-8). Thus, p34^{CDC28} seems to be phosphorylated on Y19 in cells arrested in S phase.

To determine the cell-cycle dependence of Y19 phosphorylation, we prepared extracts from wt cells arrested at Start with α -factor or arrested in mitosis with the antimicrotubule drug benomyl. No phosphotyrosine was detected on p34 isolated from G1- or M-phase cells (Fig. 1b lanes 10 and 12), suggesting that the tyrosine phosphorylation of p34^{CDC28}, like that of p34 in other organisms⁷, is restricted to the S and G2 phases of the cell cycle. This biochemical difference also indicates that drugs that inhibit S phase and M phase arrest budding yeast cells at different points in the cell cycle, although the arrest points cannot be distinguished morphologically.

We examined the phenotype of *cdc28-F*, and of another mutant, *cdc28-AF*, which has an additional substitution of threonine to alanine at position 18 (residues homologous to T18 and Y19 coordinately regulate p34^{cdc2} in vertebrate cells¹⁰). The doubling time, size, and cell cycle distribution (as assayed by fluorescence-activated cell sorting (FACS)) of *cdc28-F* and *cdc28-AF* strains were very similar to wild type, indicating that mitosis was not induced prematurely in the mutants (data not shown).

Do cells carrying *cdc28-F* and *cdc28-AF* mutations arrest normally in response to the presence of damaged DNA? Irradiation produces double-stranded breaks in DNA and, in wild-type cells, results in a cell cycle arrest at a G2 'checkpoint' that is passed only when the DNA damage has been repaired¹. In cells carrying mutations in genes, such as *RAD9*, that regulate arrest at the checkpoint, division proceeds despite the DNA damage and the cells die¹¹. We compared wt *CDC28*, *cdc28-F*, *cdc28-AF* and *rad9* cells for their resistance to X-ray irradiation (Fig. 2a). Although *rad9* cells died rapidly, the fraction of survivors was identical in the other strains, indicating that the operation of the *RAD9*-dependent checkpoint does not require phosphorylation of p34^{CDC28} on Y19.

To examine the dependence of cell division on the completion of S phase, we treated cells with sublethal doses of hydroxyurea, which delays the cell cycle by inhibiting DNA synthesis. Cells that are defective in feedback controls triggered by unreplicated DNA should die in the presence of hydroxyurea because they divide before S phase is completed. Wild-type *CDC28*, *cdc28-F* and *cdc28-AF* strains remained highly viable in hydroxyurea (Fig. 2b). As a positive control, we examined *mec1* cells (provided by T. Wienert), which do not arrest correctly in the presence of hydroxyurea, and observed that they died rapidly. Thus, the phosphorylation of p34^{CDC28} on Y19 is not essential for the correct operation of S-phase feedback control. To look

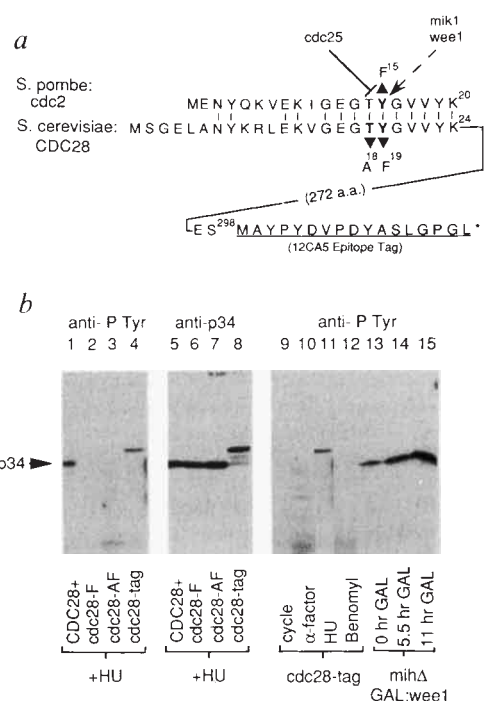
for more subtle phenotypes, we grew genetically marked *CDC28* and *cdc28-AF* strains together for 100 generations in rich medium, minimal medium and media supplemented with sublethal doses of benomyl and hydroxyurea; we found that the difference in the reproduction rates of wild-type and mutant cells was less than 0.5% per generation in all cases (data not shown).

Reduced tyrosine phosphorylation of p34^{cdc2} makes *S. pombe* cells supersensitive to hydroxyurea². Does the lack of a similar phenotype in *S. cerevisiae* cells carrying *cdc28-F* reflect a difference in the way that the *CDC28* and *cdc2* gene products respond to tyrosine phosphorylation? To investigate this, we expressed *cdc2-F15* in *S. cerevisiae* and *cdc28-F* in *S. pombe*. *S. cerevisiae* cells carrying either a wt *cdc2* gene or the *cdc2-F15* mutant remained viable in the presence of hydroxyurea (Fig. 2c). By contrast, *S. pombe* cells carrying *CDC28* grew on hydroxyurea, but those that carried *CDC28-F* did not (Fig. 2d) although the rate of killing was quite low (about 10% loss of viability per generation; data not shown). Thus, although budding and fission yeast differ in their sensitivity to the presence of p34 lacking a phosphorylatable tyrosine, this is not caused by differences between the *cdc2* and *CDC28* gene products but must reflect other physiological differences.

In fission yeast, the extent of phosphorylation of Y15 of p34^{cdc2} reflects the balance between the activity of *cdc25* (ref.

FIG. 1 Regulation of the tyrosine phosphorylation of p34^{CDC28}. a, A comparison of the amino termini of the products of the *CDC28* gene in *S. cerevisiae* and the *cdc2* gene in *S. pombe* is shown. Bold arrows, positions of mutations that were introduced by site-directed mutagenesis; in *CDC28* these are a tyrosine to phenylalanine substitution at position 19 (*cdc28-F*) and a double mutant combining F19 with a threonine to alanine mutation at position 18 (*cdc28-AF*). In *cdc2*, tyrosine 15 was mutated to phenylalanine (*cdc2-F15*). The opposing action on the phosphorylation of tyrosine 15 of p34^{cdc2} by the phosphatase encoded by the *cdc25* gene^{12,13} and the kinases encoded by *wee1* (ref. 21) and *mik1* (ref. 15) is indicated schematically. To facilitate the detection of p34^{CDC28} it was fused to a short epitope tag. A gene encoding this epitope-tagged protein seems to function normally because it could be substituted for the endogenous *CDC28* gene without affecting cell growth. The lowest line shows the sequence of the epitope tag; it was fused to the C terminus of p34^{CDC28} using site-directed mutagenesis (non-native sequences are underlined). The epitope is derived from the HA protein of the influenza virus⁹ and, when fused to p34^{CDC28}, reacts with the 12CA5 monoclonal antibody under both native and denaturing conditions. b, Immunoblots of p34 precipitates from extracts of cells carrying wild-type and mutant *CDC28* genes and arrested at various points in the cell cycle. In each panel, noncontiguous lanes from a single exposure of one filter are shown. Precipitates were prepared from samples containing equivalent amounts of total protein using p13-suc1 covalently coupled to agarose beads⁴. p13-suc1 is *S. pombe* protein that binds tightly to p34. Similar results to those shown here have been obtained by immunoprecipitation of tagged p34 with 12CA5. Lanes 1–4, samples from cells carrying wt *CDC28*, the *cdc28-F19* and *cdc28-A18F19* mutations, and epitope-tagged *CDC28* were arrested in S phase with hydroxyurea and probed with anti-phosphotyrosine antibodies; lanes 5–8, the same immunoblot reprobed with an anti-p34 serum to determine the amount of p34 in each sample; lanes 9–15, antiphosphotyrosine immunoblots of samples from cells carrying epitope-tagged *CDC28* and arrested at various points in the cell cycle (9–12) or from cells in which the *MIH1* gene had been deleted (13–15) and *S. pombe wee1* gene overexpressed from the *GAL1* promoter for 5.5 (14) or 11 h (15).

METHODS. Mid-log cultures of *MATa*, *bar1* cells were grown in rich medium. They were then arrested in G1 by the addition of α -mating factor to 2.5 $\mu\text{g ml}^{-1}$, in S phase by the addition of hydroxyurea to 15 $\mu\text{g ml}^{-1}$ and in mitosis by the addition of benomyl to 15 $\mu\text{g ml}^{-1}$. α -factor and hydroxyurea treatment were for 3 h at 30 °C, and benomyl for 3 h at 25 °C; the synchrony of the arrests was established by microscopic examination of the cells. The samples in lanes 13–15 were prepared from *MATa* cells carrying a disruption of the *MIH1* gene and several integrated copies of a construct that fuses the *GAL1/10* promoter to the *S. pombe wee1* coding sequence. *MIH1* was disrupted using a *LEU2* marker (*mih1ΔLEU2*) and the *GAL1:wee1* construct was integrated at the *URA3* locus as described previously¹⁶. These cells were grown in raffinose and the expression of the *wee1* gene induced in mid-log cultures by the addition of galactose. Extracts



were prepared by mechanical disruption using glass beads as described previously²² in breakage buffer containing 50 mM MOPS, pH 7.2, 15 mM MgCl_2 , 10 mM EGTA, 25 mM NaF, 50 mM β -glycerophosphate, 1 mM sodium orthovanadate, 1 mM PMSF, 0.5 mM *N*-tosyl-L-phenylalanine chloromethyl ketone and 1 $\mu\text{g ml}^{-1}$ each of leupeptin and chymostatin. Total protein (0.5 mg) was diluted into 0.25 ml in a buffer similar to breakage buffer but containing 50 mM Tris, pH 7.4 instead of MOPS and supplemented with 250 mM NaCl and 0.1% Triton. p13 beads (20 μl) prepared as previously described⁴ were added to each sample, incubated at 4 °C for 2 h, and washed with 50 mM Tris, pH 7.4, 150 mM NaCl, 1% Nonidet P-40, 0.05% deoxycholate, 0.1% SDS and then eluted with SDS sample buffer, run on a 15% SDS-polyacrylamide gel and transferred to nitrocellulose. Filters were blocked, incubated with undiluted culture supernatant (supplemented with Tween 20 to 0.2%) from cells expressing the 4G10 antiphosphotyrosine monoclonal antibody²³ and antibody conjugates detected by incubation with ¹²⁵I-labelled goat anti-mouse serum; filters were then stripped, reprobed with a polyclonal anti-p34^{CDC28} antibody (a gift from Ray Deshaies) and detected with alkaline phosphatase-conjugated secondary antibodies.

12), which encodes a tyrosine phosphatase^{13,14}, and a pair of functionally redundant kinases encoded by the *wee1* and *mik1* genes¹⁵. Enoch and Nurse have proposed that *cdc25* mediates the S-phase feedback control in *S. pombe*⁸. Thus, one explanation for our failure to find a phenotype associated with the *cdc28-AF* mutation in budding yeast is that p34^{CDC28} has sites of phosphorylation in addition to Y19 (and T18) that are targets of *cdc25/wee1* action. We tested this possibility by taking advantage of the observation that the simultaneous deletion of *MIH1* (a homologue of the *S. pombe cdc25* gene¹⁶) and overexpression of *wee1* arrests *S. cerevisiae* cells in a G2-like state with hyperphosphorylated p34 (Fig. 1b, lanes 13–15)¹⁶. If this arrest is due primarily to the phosphorylation of p34^{CDC28} on T18 and Y19, the *cdc28-AF* mutation should suppress the growth arrest. In agreement with this, we find almost complete suppression of the arrest by *cdc28-AF*, and somewhat weaker suppression by *cdc28-F* (Fig. 3). Moreover, suppression by *cdc28-AF* is dominant to *CDC28*, in a manner analogous to the dominance of *cdc2-F15* to *cdc2* in *S. pombe*³. Thus, the arrest imposed by the overexpression of *wee1* and deletion of *MIH1* seems to involve the hyperphosphorylation of T18 and Y19 of p34^{CDC28}.

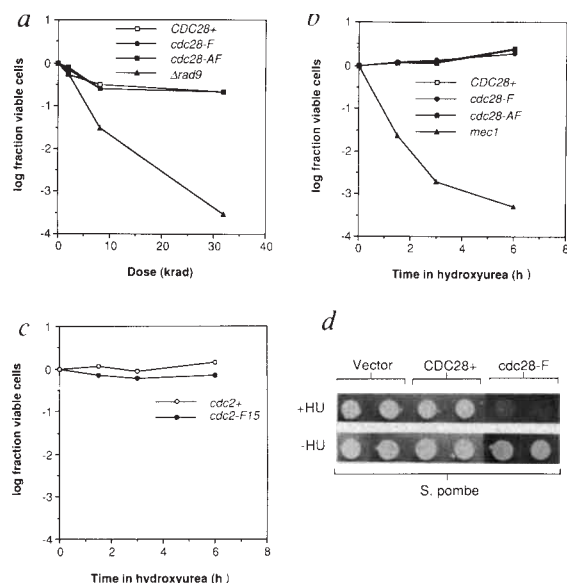


FIG. 2 Effect of p34 mutations on the viability of budding and fission yeast following DNA damage or the inhibition of DNA synthesis. **a**, Cell survival was determined for asynchronous liquid cultures of *S. cerevisiae* cells in which double-stranded breaks in DNA were introduced by exposure to X-ray irradiation. Curves are shown for strains carrying either a wt *CDC28* allele, or *cdc28-F19* or *cdc28-A18F19* alleles in place of wt *CDC28* and also for *CDC28* strains carrying a complete disruption of the *RAD9* gene¹¹ as indicated. **b**, Survival in 10 mg ml⁻¹ hydroxyurea of *S. cerevisiae* strains containing various *CDC28* alleles or the hydroxyurea-supersensitive *mec1* mutation (T. Weinert, personal communication). In the absence of hydroxyurea, cell number increased six to eightfold, demonstrating that hydroxyurea treatment substantially lengthens the cell cycle. **c**, Survival in 10 mg ml⁻¹ hydroxyurea of *S. cerevisiae* cells containing a complete disruption of the endogenous *CDC28* locus and a high-copy number (2 micron) plasmid that carries a fusion gene linking the *CDC28* promoter to the coding sequence of a wild-type *S. pombe cdc2* cDNA or of a *cdc2-F15* mutation. **d**, Growth on solid media lacking leucine of wild-type *S. pombe* cells containing a high copy number *LEU2* plasmid with no insert (vector), a wild-type *CDC28* gene (*CDC28+*), or a gene with the *cdc28-F19* mutation (*cdc28-F*) in the presence or absence of 1 mg ml⁻¹ hydroxyurea. A single exposure of non-contiguous patches of cells from a single plate is shown. **METHODS**. Following different doses of X-ray irradiation, or different lengths of treatment with hydroxyurea, cells were plated on solid media and the fraction of viable cells calculated by counting colonies after 3 days of growth at 25 °C; the viability of untreated cells at time 0 was set at 100%. The data were collected from cultures treated in parallel and each data point represents the average of two or more determinations which typically varied by less than 10%.

This suggests that the failure of the *cdc28-AF* mutation to inactivate S-phase feedback control in *S. cerevisiae* is not a consequence of the presence of additional sites of *MIH1* action on p34^{CDC28}.

It is thought that high levels of tyrosine phosphorylation of p34 in G2 inhibit the kinase⁷ and, in cells containing unreplicated DNA, arrest division by eliminating p34 activity⁸. To investigate this in budding yeast, we measured p34^{CDC28} activity, using histone H1 as a substrate, in extracts from cells arrested at various points in the cell cycle. We found that p34 activity is extremely low in cells arrested at Start with α -factor, increases at least 20-fold in cells arrested in S phase with hydroxyurea and rises a further three-fold in cells arrested in mitosis with benomyl (Fig. 4a). Similar regulation is observed in *cdc* mutants¹⁷ arrested at the non-permissive temperature (Fig. 4b). Thus, the activity of p34^{CDC28} seems to be high during S phase even though Y19 phosphorylation is also maximal. Arrest caused by unreplicated DNA does not seem to involve the elimination of p34^{CDC28} activity, at least as measured *in vitro*. Mutations that prevent the tyrosine phosphorylation of p34^{CDC28} do not increase kinase activity in S-phase arrested cells (Fig. 4c); conversely, conditions that induce the hyperphosphorylation of p34 and arrest cells in G2 (ref. 16) (the deletion of *MIH1* and overexpression of *wee1*, Fig. 3) do not seem to reduce kinase activity below that found in S phase (Fig. 4c). These data suggest that in *S. cerevisiae*, in contrast to what has been proposed previously⁷, Y19 phosphorylation may regulate cell-cycle progression by a mechanism that does not simply involve the elimination of the catalytic activity of p34.

We have shown that p34^{CDC28} is phosphorylated in S phase on Y19, a site that is homologous to the tyrosine modified in other p34 proteins and that Y19, and possibly an adjacent threonine, are the principal sites of action of *MIH1*. But mutations that prevent the phosphorylation of p34^{CDC28} on T18 and Y19 do not accelerate mitosis and do not disrupt the feedback controls that link cell division with the completion of DNA synthesis. Despite the ubiquity in eucaryotes of the tyrosine phosphorylation of p34 (ref. 7), and the conservation, even in budding yeast, of the *wee1/cdc25* pathway¹⁶, tyrosine phosphorylation of p34 does not seem to constitute the universal rate-limiting regulator of the entry into M phase.

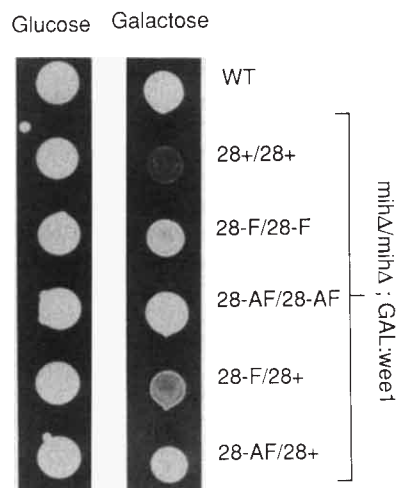


FIG. 3 Effect of *CDC28* mutations on the G2-like arrest of budding yeast cells lacking *MIH1* and overexpressing *wee1*. Growth on solid media of diploid strains that are wild-type (WT) or that are homozygous for a disruption of the *mih1* gene, that overexpress the *S. pombe wee1* protein from an integrated *GAL:wee1* fusion construct (present in several linked copies; see Fig. 1) and that carry various combinations of *CDC28* alleles. 28+, Wild type *CDC28* locus; 28-F, *cdc28-F19* mutation; 28-AF, *cdc28-A18F19* mutation. *Wee1* expression is strongly induced on plates containing galactose and is repressed on plates containing glucose; both types of plates also contained 1 mM sodium orthovanadate.

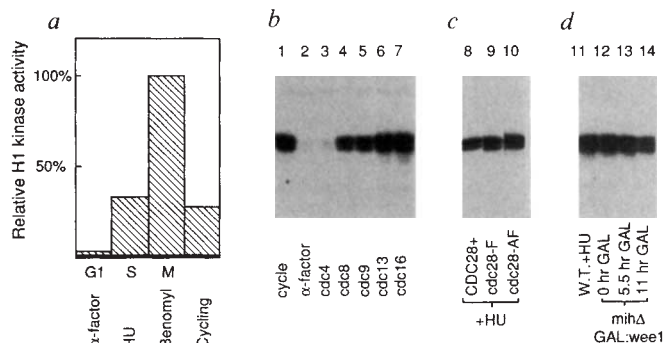


FIG. 4 Kinase activity of p34^{CDC28} in cells arrested at various stages in the cell cycle. **a**, The kinase activity of p34^{CDC28} in cells arrested in G1 and S phase is shown as a percentage of the activity in mitosis (defined as 100%). Cultures of cells carrying tagged *CDC28* were arrested with α -factor, hydroxyurea and benomyl, and extracts prepared as described in Fig. 1; p34 was immunoprecipitated with 12CA5 antibodies and kinase activity measured using histone H1 as a substrate essentially as described²². The solid bar (1.5%) represents the background level of activity in precipitates from cell extracts that contain untagged p34^{CDC28}. Each data point represents an average of three or more determinations done at different extract concentrations. **b**, The kinase activity of p34^{CDC28} precipitated with p13 beads from extracts of cells carrying temperature-sensitive mutations in genes required for progression through the cell cycle (*cdc* genes¹⁷) after arrest at 37 °C for 3 h *cdc4* prevents the initiation of DNA synthesis, *cdc8* blocks DNA synthesis, *cdc9* is defective in DNA ligase, *cdc13* arrests cells in G2, and *cdc16* arrests cells in mitosis. Analysis of these precipitates, and those of **c** below, by immunoblotting with anti-p34^{CDC28} antiserum shows that they contain equivalent amounts of p34. Each panel shows samples assayed in parallel; the activities of samples in different panels cannot be compared directly. **c**, p34^{CDC28} kinase activity in p13 precipitates from extracts of cells carrying various *CDC28* alleles and arrested with hydroxyurea (lanes 8–11), or **d**, of cells carrying the *mih1ΔLEU2* mutation and the *GAL1::wee1* fusion construct (Fig. 1) in conditions in which *wee1* expression is repressed (lane 12) or has been induced for 5.5 h (lane 13) or 11 h (lane 14). The extracts used in lanes 12–14 are identical to those in lanes 12–15 of Fig. 1.

Why does interfering with the phosphorylation of p34 on tyrosine have different effects on feedback control in fission and budding yeasts? Perhaps there exist several parallel controls (one of which may be tyrosine phosphorylation of p34) whose relative importance varies between the two organisms. Alternatively, it is possible that S-phase feedback control in *S. pombe* operates primarily through a mechanism other than the regulation of Y15 phosphorylation, but that *cdc2-F15* acts to overwhelm this restraining mechanism by strongly accelerating the entry into mitosis. The observation that the overproduction of *string* (a homologue of *S. pombe cdc25*) in *Drosophila* does not induce mitosis in cells containing unreplicated DNA¹⁸, suggests that other organisms may not require the tyrosine phosphorylation of p34 for feedback control.

We have demonstrated the existence of a mechanism independent of the tyrosine phosphorylation of p34^{CDC28} by which *S. cerevisiae* cells that contain unreplicated or damaged DNA arrest their division and regulate the entry into mitosis. One possibility is that this mechanism requires an uncharacterized modification of the maturation promoting factor (MPF) complex. But two pieces of evidence suggest that it does not simply involve a second inhibitory modification of p34^{CDC28}. First, in a screen of 8,000 heavily mutagenized plasmids that carry *cdc28-F*, we were unable to identify any hydroxyurea-sensitive *CDC28* alleles (L. Godley and P.K.S., unpublished observations). Second, T18 and Y19 are the only residues in p34^{CDC28} whose phosphorylation rises during S phase¹⁹. A second possibility is that there exists a regulator of mitosis other than MPF and that this regulator is rate limiting in *S. cerevisiae* and is the primary target of feedback controls. The *nimA* kinase of *Aspergillus nidulans*, which appears to be unrelated to the

MPF kinase but is a key activator of mitosis, may be such a regulator²⁰. □

Received 24 October; accepted 14 November 1991.

- Hartwell, L. H. & Weinert, T. A. *Science* **246**, 629–634 (1989).
- Enoch, T. & Nurse, P. *Cell* **60**, 665–673 (1990).
- Gould, K. L. & Nurse, P. *Nature* **342**, 39–45 (1989).
- Solomon, M. J., Glotzer, M., Lee, T. H., Philippe, M. & Kirschner, M. W. *Cell* **63**, 1013–1024 (1990).
- Krek, W. & Nigg, E. A. *EMBO J.* **10**, 305–316 (1991).
- Dracetta, G. et al. *Nature* **336**, 738–744 (1988).
- Nurse, P. *Nature* **344**, 503–508 (1990).
- Enoch, T. & Nurse, P. *Cell* **65**, 921–923 (1991).
- Field, J. et al. *Molec. cell. Biol.* **8**, 2159–2165 (1988).
- Krek, W. & Nigg, E. A. *EMBO J.* **10**, 3331–3341 (1991).
- Weinert, T. A. & Hartwell, L. H. *Science* **241**, 317–322 (1988).
- Russell, P. & Nurse, P. *Cell* **45**, 145–153 (1986).
- Dunphy, W. G. & Kumagai, A. *Cell* **67**, 189–196 (1991).
- Gautier, J., Solomon, M. J., Boohar, R. N., Bazan, J. F. & Kirschner, M. W. *Cell* **67**, 197–211 (1991).
- Lundgren, K. et al. *Cell* **64**, 1111–1122 (1991).
- Russell, P., Moreno, S. & Reed, S. I. *Cell* **57**, 295–303 (1989).
- Hartwell, L. H. *Bact. Rev.* **38**, 164–198 (1974).
- Edgar, B. A. & O'Farrell, P. H. *Cell* **62**, 469–480 (1990).
- Amon, A., Surana, U., Muroff, I. & Nasmyth, K. *Nature* **355**, 368–371 (1992).
- Osmani, A. H., McGuire, S. L. & Osmani, S. A. *Cell* **67**, 283–291 (1991).
- Russell, P. & Nurse, P. *Cell* **49**, 559–567 (1987).
- Surana, U. et al. *Cell* **65**, 145–161 (1991).
- Druker, B. J., Mamon, H. J. & Roberts, T. M. *New Engl. J. Med.* **321**, 1383–1391 (1989).

ACKNOWLEDGEMENTS. We thank H. Varmus for encouragement and help, R. Deshaies for help, T. Weinert, R. Li, J. Ogas, R. Boohar and D. Morgan for reagents, Q. P. Li for technical assistance, A. Amon and K. Nasmyth for the communication of results before publication, members of the UCSF cell-cycle group for discussions, and the IMP in Vienna for its hospitality. P.K.S. is a Fellow of the Jane Coffin Childs Memorial Fund and P.K.S. and A.W.M. are biomedical scholars of the Lucille P. Markey Charitable Trust; this work was supported by these organizations as well as by grants from the NIH to A.W.M. and H. Varmus.

Regulation of p34^{CDC28} tyrosine phosphorylation is not required for entry into mitosis in *S. cerevisiae*

Angelika Amon, Uttam Surana, Ivor Muroff & Kim Nasmyth*

Institute of Molecular Pathology, Dr. Bohr Gasse 7, A-1030 Vienna, Austria

PROGRESSION from G2 to M phase in eukaryotes requires activation of a protein kinase composed of p34^{cdc2/CDC28} associated with G1-specific cyclins (reviewed in ref. 1). In some organisms the activation of the kinase at the G2/M boundary is due to dephosphorylation of a highly conserved tyrosine residue at position 15 (Y15) of the *cdc2* protein^{2–6}. Here we report that in the budding yeast *Saccharomyces cerevisiae*, p34^{CDC28} also undergoes cell-cycle regulated dephosphorylation on an equivalent tyrosine residue (Y19). However, in contrast to previous observations in *S. pombe*⁶, *Xenopus*^{2,3} and mammalian cells^{4,5}, dephosphorylation of Y19 is not required for the activation of the CDC28/cyclin kinase. Furthermore, mutation of this tyrosine residue does not affect dependence of mitosis on DNA synthesis nor does it abolish G2 arrest induced by DNA damage. Our data imply that regulated phosphorylation of this tyrosine residue is not the 'universal' means by which the onset of mitosis is determined. We propose that there are other unidentified controls that regulate entry into mitosis.

To investigate whether entry into mitosis in *S. cerevisiae* is associated with dephosphorylation of p34^{CDC28} we compared the tryptic phosphopeptide maps of p34^{CDC28} in various *cdc* mutants arrested at different stages of the cell cycle. Among the phosphopeptides derived from p34^{CDC28}, phosphorylation of only one major peptide (peptide 1) is cell-cycle regulated (Fig. 1). Although peptide 1 is absent in cells arrested in G1 by

* To whom correspondence should be addressed.

pheromone (data not shown) or shortly after Start by the *cdc34-2* mutation (Fig. 1a), it is present in mutants arrested in S phase (*cdc7-1* and *cdc8-1*; Fig. 3b) or G2 (*cdc9-1* and *cdc13-1*; Figs 1b and 3b). Peptide 1 is absent in *cdc23-1* mutants (Fig. 1c) that have a terminal phenotype similar to that of *cdc13-1* mutants (see legend to Fig. 1) and in cells blocked in mitosis by the anti-microtubule agent nocodazole (Fig. 3b). Phosphoamino-acid analysis of this peptide showed that it contains phosphotyrosine (Fig. 1d). We assume that the phosphorylation is on tyrosine 19 (Y19) because replacement by phenylalanine (F19) not only eliminated peptide 1 from the phosphopeptide

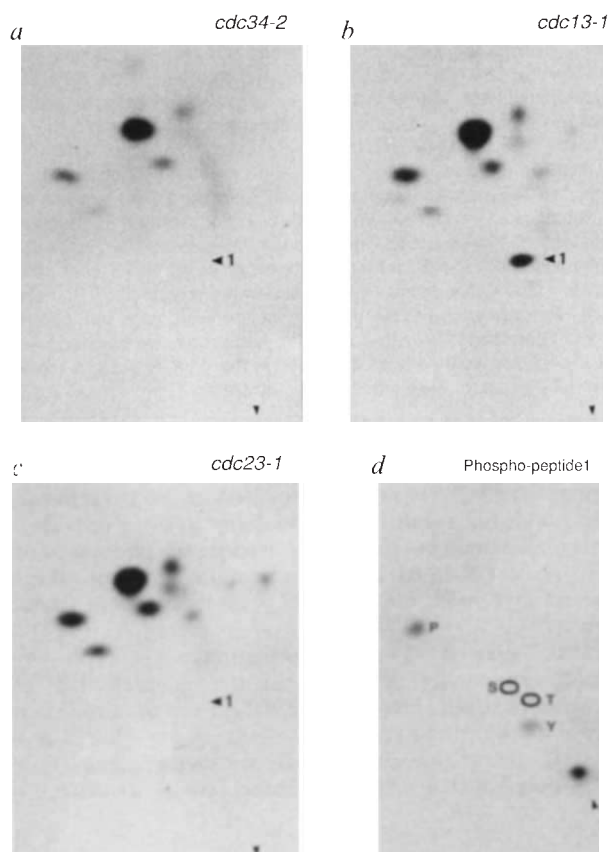


FIG. 1 Y19 phosphorylation is cell-cycle regulated. *a*, Tryptic phosphopeptides of p34^{CDC28} isolated from *cdc34-2* (K1983) cells arrested shortly after Start. *b*, *cdc13-1* (K2034) cells arrested in G2. *c*, *cdc23-1* (K2531) cells probably arrested in mitosis. (1, indicates position of phosphopeptide 1.) *d*, Phosphoamino-acid analysis of phosphopeptide 1. (S, serine; T, threonine; Y, tyrosine; P, free phosphate.) Note that the difficulty in distinguishing G2 from M-phase arrest by cytological means in *S. cerevisiae* may now be overcome by measuring Y19 phosphorylation. By the criterion of Y19 phosphorylation *cdc13-1* mutants arrest in G2 whereas *cdc23-1* mutants arrest in metaphase.

METHODS. All yeast strains used in this and subsequent experiments were congenic with or at least three times backcrossed to W303 (*Mata ade2-1 leu2-3 trp1-1, his3-11, 15 can1-100 ura3 GAL psi +*). *cdc* mutants (40 ml; absorbance at 600 nm, 0.8) were grown in phosphate-depleted YEPD medium as described previously¹⁸, arrested for 3 h at 37 °C, and labelled for 10 min with 0.1 mCi ml⁻¹ [³²P]orthophosphate. Immunoprecipitations were done as described¹⁹, with a 1:150 dilution of polyclonal antiserum (R49) raised against the CDC28 protein expressed in *Escherichia coli*. p34 immunoprecipitates were subjected to tryptic digestion; 900 c.p.m. (Cerenkov) were loaded onto thin-layer chromatography (TLC) plates (Merck). The tryptic peptides were separated in two dimensions as described¹⁹. Electrophoresis was at pH 8.9 in the horizontal dimension for 80 min at 1 kV with the anode on the left. Exposure time was 6 days at -70 °C with an intensifying screen. Phosphopeptide 1 was eluted from the TLC plate and processed as described¹⁹ followed by hydrolysis with HCl and separated by two-dimensional thin layer electrophoresis as described²⁰. Arrows at the bottom indicate loading origin.

map (Fig. 2), but also eliminated all phosphotyrosine from p34^{CDC28} (data not shown). Our data are consistent with the pattern of Y15 phosphorylation found in other organisms^{6,7}, that is, high phosphorylation in S and G2 and low in M phase.

Surprisingly, the substitution of Y19 by F19 has no noticeable effect on the rate of cell division, cell size or the length of the G2 phase (data not shown), although an equivalent mutation in *S. pombe* causes cells to enter mitosis prematurely⁶. In chicken cells phosphorylation of both Y15 and T14 is important for inhibition of p34^{cdc2} (ref. 8). T18 (the equivalent of T14 in *cdc2*) may also be very weakly phosphorylated in *S. cerevisiae* (see legend to Fig. 2). However, we found that the growth characteristics of *cdc28A18F19* double mutants are similar to those of *cdc28F19* single mutants (data not shown). Furthermore, CDC28-dependent histone H1 kinase associated with the G2-specific cyclin CLB2 is unaffected by the F19 mutation in cycling cells or in cells arrested in mitosis (nocodazole) and in G2 (*cdc13-1*) (Fig. 3a).

To avoid initiation of mitosis before completion of DNA replication, specific 'checkpoint' controls exist that can detect unreplicated or damaged DNA and signal arrest in G2 (reviewed in ref. 9). One such control is imposed by the *RAD9* gene in *S. cerevisiae* which causes G2 arrest in the event of DNA damage¹⁰. Work in *S. pombe* implies that an equivalent type of G2-arrest in this species is due to inhibition of the *cdc2* kinase activity by phosphorylation of Y15 in *cdc2* (refs 11-14). To determine whether Y19 is the target of DNA-checkpoint controls in *S. cerevisiae*, we analysed the effects of the F19 mutation on the cell-cycle arrest of *cdc8-1* mutants that arrest in S phase¹⁵ and *cdc13-1* mutants that show a *RAD9*-dependent G2 arrest¹⁰. Arrest of neither mutant was altered by the *cdc28F19* mutation; both the *cdc8-1 cdc28F19* and the *cdc13-1 cdc28F19* double mutants arrest uniformly as uninucleate budded cells containing

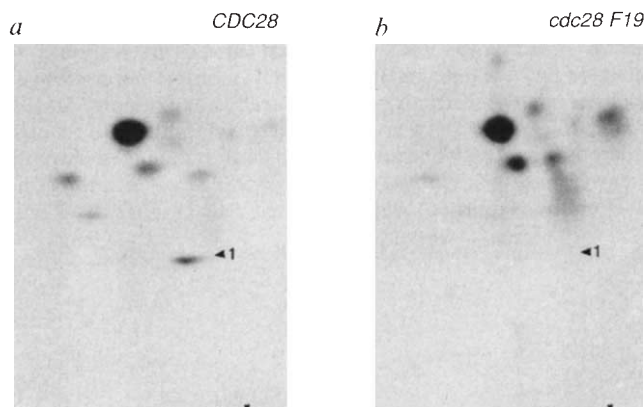
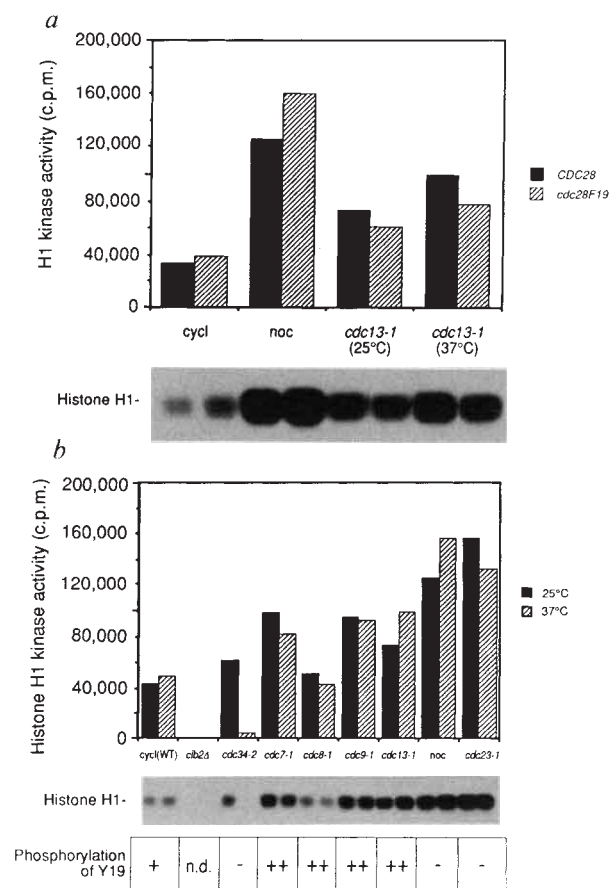


FIG. 2 Comparison of tryptic phosphopeptides of p34^{CDC28} and p34^{cdc28F19}. *a*, Tryptic phosphopeptides of p34^{CDC28} from exponentially growing wild-type cells (K1534). *b*, Exponentially growing *cdc28F19* cells (A377). We also detected a very minor peptide (regulated similarly to peptide 1, although not visible here) that contains both phosphorylated threonine and tyrosine. This peptide may contain both phosphorylated T18 and Y19. **METHODS.** The Y19 and the T18 residues were replaced by phenylalanine and alanine, respectively, using the oligonucleotide-directed *in vitro* mutagenesis system manufactured by Amersham. The *cdc28A18F19* allele was cloned into the yeast centromere vector YCplac22 (ref. 21) and transformed into a diploid heterozygous for a *cdc28:URA3* deletion (A372). A *cdc28:URA3 Ycplac22-cdc28A18F19* segregant (A442) was obtained and analysed further. The *cdc28F19* mutant allele was cloned into the yeast integration vector Ylplac204 (ref. 21) and transformed into A372. Transformants were obtained where the plasmid had integrated at the *CDC28* locus (A373) (confirmed by Southern analysis). Dissection of this diploid produced a segregant (A377) that had the *CDC28* locus replaced by the *cdc28F19* mutant allele. Cells bearing the wild-type *CDC28* gene (K1534) or the *cdc28F19* mutation (A377) were grown to log phase at 30 °C before labelling with ³²P. Labelling, immunoprecipitations and separation of tryptic phosphopeptides were done as described in the legend to Fig. 1; 700 c.p.m. were loaded on TLC plates; exposure time was 8 days.



a pre-anaphase mitotic spindle (data not shown). A failure of control mechanisms to signal arrest in G2 in response to damaged DNA is reflected in loss of cell viability after prolonged incubation at the restrictive temperature¹⁰. For example, *cdc13-1 rad9::URA3* cells show a rapid loss of viability due to cells undergoing successive rounds of mitosis with damaged chromosomes (Fig. 4b)¹⁰. We find that the *cdc28F19* mutation has no effect on the viability of *cdc8-1*-arrested cells (Fig. 4a) and only

FIG. 3 CLB2-dependent histone H1 kinase activity and Y19 phosphorylation. **a**, Comparison of CLB2-dependent histone H1 kinase in *CDC28* (K1534), *cdc28F19* (A377) strains grown at 25 °C in the presence (noc) or absence (cycl) of nocodazole (15 $\mu\text{g ml}^{-1}$), in *cdc13-1 CDC28* (K2034) and in *cdc13-1 cdc28F19* grown either at 25 °C or 37 °C. **b**, CLB2-dependent histone H1 kinase analysed at 25 °C and 37 °C in wild-type and various *cdc* mutants (wild-type (K1534), *clb2::LEU2* (K1890), *cdc34-2*(K1983), *cdc7-1*(K2032), *cdc8-1*(K2536), *cdc9-1*(K2538), *cdc13-1* (K2034), wild-type (K1534) arrested with nocodazole (15 $\mu\text{g ml}^{-1}$) and *cdc23-1* (K2531)). The table below the bar chart shows the state of tyrosine19 (Y19) phosphorylation in the various *cdc* strains grown at 37 °C: -, +, ++ indicates absence, presence of Y19 phosphorylation to the extent observed in exponentially growing wild-type cells and enhanced Y19 phosphorylation, respectively. n.d., Not determined.

METHODS. Cells were grown as indicated, collected and extracts were prepared by breaking cells with glass beads in lysis buffer (50 mM Tris pH 7.5 containing 1% Triton X-100, 1% deoxycholate and 0.1% SDS, 60 mM β -glycerophosphate, 40 mg l^{-1} aprotinin, 0.1 mM sodium orthovanadate and 15 mM *p*-nitrophenyl phosphate). Immunoprecipitations were done as described²² with minor modifications. CLB2-associated kinase was immunoprecipitated with CLB2-specific antibodies (at dilution 1:40) raised against the CLB2 protein expressed in *E. coli*. The kinase activity in the immunoprecipitate was measured at 25 °C and quantitated as described²³. The histone H1 kinase activity measured in these assays is *CDC28*-dependent because *cdc28-4* (K1989) mutants either grown at 25 °C or at 37 °C completely lack kinase activity (kinase assays were done at 25 °C; data not shown). Appearance of Y19 phosphorylation was analysed by tryptic phosphopeptide maps. Growth conditions, immunoprecipitation and separation of peptides were done as described in the legend to Fig. 1. Between 700 and 1,000 c.p.m. were loaded onto TLC plates.

a very modest effect on that of *cdc13-1* mutants (Fig. 4b). The failure of the *cdc28F19* mutation to alleviate G2 arrest in *cdc13-1* mutants does not result from its inability to drive cells through mitosis under these circumstances because the phenotype of the *cdc13-1 rad9::URA3 cdc28F19* triple mutant is identical to that of the *cdc13-1 rad9::URA3 CDC28* double mutant (data not shown).

That the state of Y19 phosphorylation is not important for ensuring dependency of mitosis on the completion of DNA replication raises the possibility that this form of mitotic control does not act by modulating the activity of the *CDC28* protein kinase. To investigate this further, we measured the level of *CDC28*-dependent histone H1 kinase activity associated with

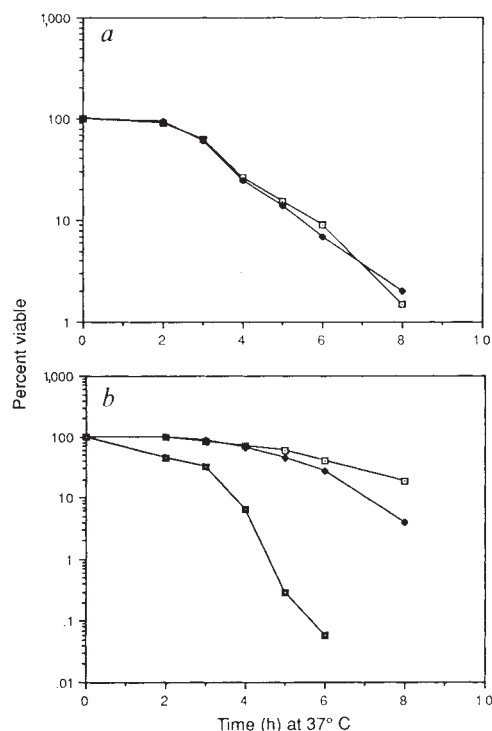


FIG. 4 Viability curves of *cdc* mutants bearing either the wild-type *CDC28* gene or the *cdc28F19* mutant gene. **a**, Viability of *cdc8-1 CDC28* (K2536) mutants ($\square$) and *cdc8-1, cdc28F19* (A404) mutants ($\blacklozenge$). **b**, Viability of *cdc13-1 CDC28* (K2034) mutants ($\square$), of *cdc13-1 cdc28F19* (A383) mutants ($\blacklozenge$) and of *cdc13-1 rad9::URA3 CDC28* (K2554) mutants ($\blacksquare$).

METHODS. *Cdc* strains carrying either the wild-type *CDC28* or the *cdc28F19* allele integrated at the *CDC28* locus were grown to stationary phase on YEPD plates at 25 °C for 36 h followed by inoculation into YEPD medium at 37 °C (10^7 cells ml^{-1}). Samples were withdrawn every hour for 8 h and plated on YEPD plates at appropriate dilutions. Plates were incubated at 25 °C and colonies were counted after 4 days.

CLB2 in various *cdc* mutants arrested in different stages of the cell cycle. There was little activity in *cdc34-2* mutants, substantial activity in cells blocked in S phase due to *cdc7-1* and *cdc8-1*, and high levels in cells arrested in G2 by the *cdc9-1* or *cdc13-1* mutations (Fig. 3b). It is remarkable that *cdc13-1* mutants arrest in G2 with a high kinase activity that is only about 50% lower than that seen in cells blocked in mitosis (that is, cells arrested by nocodazole or the *cdc23-1* mutation). Although it is difficult to decide whether the small difference in kinase activity between *cdc13-1* and *cdc23-1* mutants is of physiological significance, our results are consistent with the notion that G2 arrest in response to unreplicated or damaged DNA may not be caused by a reduction in CDC28 kinase activity.

It is currently thought that activation of a G2 specific form of the *cdc2/CDC28* protein kinase is sufficient to derive cells into M phase^{12,16,17} and that dependence of mitosis on the completion of DNA synthesis is due to inhibition of *cdc25*-mediated dephosphorylation of a *cdc2* tyrosine residue (Y15 in *S. pombe* and Y19 in *S. cerevisiae*)¹². Our results bring into question the universality of this model. We note that chicken *p34^{cdc2}* mutated at both T14 and Y15 is also not capable of driving cells into, let alone through, mitosis in human cells⁸. Thus there must be control mechanisms other than Y19 regulation that ensure the correct timing of mitosis. The budding yeast provides an opportunity to uncover such novel regulatory pathways that may have targets other than CDC28. □

Received 29 October; accepted 14 November 1991.

1. Pines, J. & Hunter, T. *New Biol.* **2**, 389–401 (1990).
2. Dunphy, W. G. & Newport, J. W. *Cell* **58**, 181–191 (1989).
3. Gautier, J., Matsukawa, T., Nurse, P. & Maller, J. *Nature* **339**, 626–629 (1989).
4. Draetta, G. & Beach, D. *Cell* **54**, 17–26 (1988).
5. Morla, A. O., Draetta, G., Beach, D. & Wang, J. Y. *J. Cell* **58**, 193–203 (1989).
6. Gould, K. L. & Nurse, P. *Nature* **342**, 39–45 (1989).
7. Krek, W. & Nigg, E. A. *EMBO J.* **10**, 305–316 (1991).
8. Krek, W. & Nigg, E. A. *EMBO J.* **10**, 3331–3341 (1991).
9. Hartwell, L. & Weinert, T. *Science* **246**, 629–634 (1989).
10. Weinert, T. & Hartwell, L. *Science* **241**, 317–322 (1988).
11. Enoch, T. & Nurse, P. *Cell* **60**, 665–673 (1990).
12. Nurse, P. *Nature* **344**, 503–508 (1990).
13. Dunphy, W. G. & Kumagai, A. *Cell* **67**, 189–196 (1991).

14. Gautier, J., Solomon, M. J., Booher, R. N., Bazan, J. F. & Kirschner, M. *Cell* **67**, 197–211 (1991).
15. Hartwell, L. *J. Bact.* **115**, 966–974 (1973).
16. Dunphy, W. & Newport, J. *Cell* **55**, 925–928 (1988).
17. Murray, A. W. & Kirschner, M. W. *Science* **246**, 614–621 (1989).
18. Rubin, G. M. *Meth. Cell Biol.* **12**, 45–64 (1975).
19. Moll, T., Tebb, G., Surana, U., Roberts, H. & Nasmyth, K. *Cell* **66**, 743–758 (1991).
20. Cooper, J. A., Sefton, B. M. & Hunter, T. *Meth. Enzym.* **9**, 387–402 (1983).
21. Gietz, R. D. & Sugino, A. *Gene* **74**, 527–534 (1988).
22. Reed, S. I., Hadwiger, J. A. & Lorincz, A. T. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4055–4059 (1985).
23. Surana, U. *et al.* *Cell* **65**, 145–161 (1991).

ACKNOWLEDGEMENTS. We thank G. Schaffner for synthesizing oligos and H. Tkadletz for figures. P. Sorger and E. Nigg for communication of results before publication and L. Ballou for reading the manuscript. This work was supported by the Austrian Industrial Research Promotion Fund.

A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene

A. E. Eriksson*, W. A. Baase, J. A. Wozniak & B. W. Matthews†

Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, University of Oregon, Eugene, Oregon 97403, USA

THE hydrophobic cores of proteins are generally well packed, with few cavities^{1,2}. Mutations in which a bulky buried residue such as leucine or phenylalanine is replaced with a small residue such as alanine can create cavities in the core of a protein (our unpublished results). The sizes and shapes of such cavities can vary substantially depending on factors such as local geometry, whether or not a cavity already exists at the site of substitution, and the degree to which the protein structure relaxes to occupy the space vacated by the substituted residue. We show by crystallographic and thermodynamic analysis that the cavity created by the replacement Leu 99 → Ala in T4 lysozyme is large enough to bind benzene and that ligand binding increases the melting temperature of the protein by 6.0 °C at pH 3.0. Benzene does not, however, bind to the cavity created by the Phe 153 → Ala replacement. The results show that cavities can be engineered in proteins and suggest that such cavities might be tailored to bind specific ligands. The binding of benzene at an internal site 7 Å from the molecular surface also illustrates the dynamic nature of proteins, even in crystals.

The two mutations, Leu 99 → Ala (L99A) and Phe 153 → Ala (F153A) in T4 lysozyme provide somewhat contrasting examples of 'cavity-creating' replacements. In pseudo wild-type T4 lysozyme (WT* in Table 1) Leu 99 and Phe 153 are completely buried in the hydrophobic core of the carboxy-terminal domain and are located next to a cavity of volume 33 Å³. When Leu 99 is replaced with Ala, the protein structure changes very little

(unpublished results) forming a cavity of volume 150 Å³ in the mutant structure. By contrast, when Phe 153 is replaced with Ala, atoms in the vicinity of the replacement move up to 1 Å, forming a smaller cavity (101 Å³). In the structure of the double mutant, L99A/F153A, a single cavity of volume 222 Å³ is formed. Cavity formation entails a substantial decrease (~2–8 kcal mol⁻¹ in different cases) in the stability of the folded protein (A.E.E. *et al.*, manuscript in preparation). These mutant proteins provided a range of structural contexts in which to explore the possibility of binding a hydrophobic molecule in the core of the protein.

There is a precedent for the binding of simple nonpolar ligands to proteins^{3–5}. In the present case, benzene was chosen as a hydrophobic probe and as the simplest phenylalanine analogue.

TABLE 1 Effect of benzene on the stability of mutant lysozymes

Protein	Concentration of benzene (mM)	pH	<i>T_m</i> (°C)	Δ <i>T_m</i> (°C)	Δ <i>H</i> (kcal mol ⁻¹)	ΔΔ <i>G</i> (kcal mol ⁻¹)
WT*	0.0	3.01	51.6	—	119.0	—
	10.4	3.01	51.5	-0.1	113.3	0.0
L99A	0.0	3.01	36.2	—	82.4	—
	2.6	3.01	40.7	4.5	97.2	1.4
	5.6	3.01	41.6	5.4	100.6	1.7
	7.5	3.01	42.2	6.0	98.8	1.9
F153A	0.0	3.01	39.5	—	72.5	—
	3.0	3.01	39.5	0.0	71.6	0.0
	7.2	3.01	39.4	-0.1	70.6	0.0
L99A/F153A	0.0	4.46	41.7	—	45.3	—
	10.4	4.46	44.7	3.0	63.5	0.6

WT* is a pseudo wild-type lysozyme in which the two cysteines present in normal wild-type lysozyme were replaced with threonine and alanine (C54T/C97A)²¹. All mutants were constructed in this background. *T_m* is the melting temperature of the protein; Δ*T_m* is the change in melting temperature of the same protein in the presence of benzene. Δ*H* is the enthalpy of unfolding at *T_m*. ΔΔ*G* is the change in the free energy of unfolding of the protein caused by the presence of benzene, estimated from the relationship²² ΔΔ*G* = Δ*S* · Δ*T_m*, where Δ*S* is the entropy of unfolding of the benzene-protein complex. A positive value of ΔΔ*G* indicates an increase in stability. The estimated error in *T_m* is ±0.2 °C and in Δ*H* is ±4 kcal mol⁻¹. Thermal unfolding²³ was done in 25 mM KCl, 20 mM KPO₄ in the presence of varying amounts of benzene. Experiments for the double mutant were at pH 4.46 as this protein is partially unfolded at pH 3.0. Because of this change in pH, the *T_m* of the double mutant cannot be directly compared with the other proteins. Folding was >95% reversible in all cases.

* Present address: Department of Molecular Biology, Biomedical Centre, Box 590, S-751 24 Uppsala, Sweden.

† To whom correspondence should be addressed.

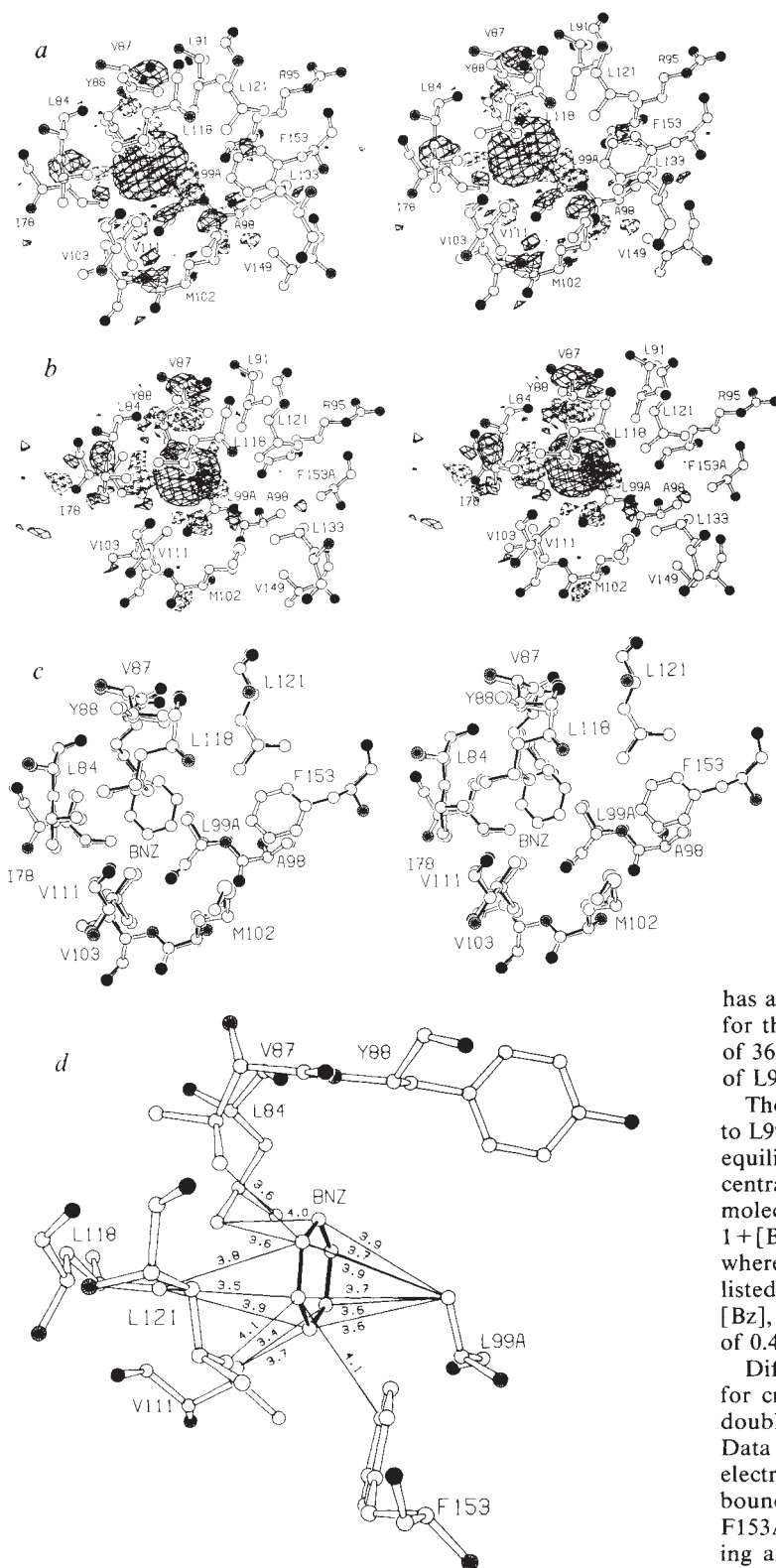


FIG. 1 *a*, Map showing the difference in electron density between benzene-soaked L99A mutant lysozyme crystals and the same mutant in the absence of benzene. Amplitudes ($F_{\text{obs,Mut+Bz}} - F_{\text{obs,Mut}}$) and phases calculated from the refined structure of the mutant lysozyme. The density is contoured at $+3\sigma$ (solid lines) and -3σ (broken lines) where σ is the r.m.s. density throughout the unit cell. Resolution 1.7 Å. In the superimposed mutant structure oxygen atoms are drawn solid, nitrogen atoms half-solid and carbon atoms as open spheres. *b*, Map showing the difference between the double mutant L99A/F153A in the presence and absence of benzene. Resolution 1.9 Å. All conventions as in *a*. The lack of density in the vicinity of residue 153 shows that benzene binds to L99A/F153A in essentially the same manner as it does to L99A. *c*, Superposition of the benzene-containing mutant lysozyme structure L99A (solid bonds) on the mutant structure in the absence of ligand (open bonds). *d*, van der Waals' contacts of benzene bound to mutant L99A; all interatomic distances are shown within an arbitrary cut-off value of 4.1 Å.

The presence of benzene increases the melting temperature (T_m), as well as enthalpies of unfolding at the melting temperature, for L99A as well as L99A/F153A (Table 1). For L99A the melting temperature increases by up to 6.0 °C. There was little or no detectable influence of benzene on the T_m of F153A or WT*, suggesting that benzene does not interact with either the folded or the unfolded forms of these proteins. (In principle benzene might interact equally with both forms, but this is discounted by the crystallographic data described below.) Because F153A

has a T_m in buffer alone of 39.5 °C, it serves as a direct control for the L99A binding data which span the temperature range of 36.2 (the T_m of the L99A in buffer alone) to 42.2 °C (the T_m of L99A in 7.5 mM benzene).

The isothermal dissociation constant, K_d , for benzene bound to L99A in this buffer was determined from a fit of the apparent equilibrium constant for thermal unfolding, K_{app} , to the concentration of benzene. With the assumption of binding of one molecule of benzene to only the folded protein, $(1/K_{\text{app}}) = 1 + [\text{Bz}]/K_d$ at the T_m of the protein in ligand-free buffer^{6,7} where [Bz] is the concentration of benzene. The data for L99A listed in Table 1 give a straight-line plot of $(1/K_{\text{app}})$ against [Bz], indicating single-site binding with a dissociation constant of 0.40 mM for the L99A-benzene complex.

Diffraction data in the presence of benzene were collected for crystals of the two single mutants, L99A and F153A, the double mutant L99A/F153A and wild-type lysozyme (WT*). Data collection statistics are summarized in Table 2. Difference electron-density maps (Fig. 1*a, b*) showed that benzene was bound to the L99A and L99A/F153A structures, but not to F153A or to wild-type lysozyme. The distinction between binding and non-binding is very clear. In the double mutant, for example (Fig. 1*b*), the electron density showing the binding of benzene at the Leu 99 site is of magnitude 8 σ , where σ is the r.m.s. density throughout the unit cell. By contrast, there is no difference density feature in the vicinity of the cavity created by the Phe 153 $\rightarrow$ Ala replacement. In the two cases where benzene was bound to the protein the structures were refined^{8,9}. Statistics are included in Table 2.

In both the L99A and L99A/F153A structures, benzene binds in a very similar manner at the site occupied by Leu 99 in wild-type lysozyme (Fig. 1*c*). The structural adjustments associated with the binding of benzene are modest. In each case the

TABLE 2 X-ray data collection and refinement statistics

Protein	WT*	WT* +Bnz	F153A +Bnz	L99A +Bnz	F153A/ L99A +Bnz
Method	Osc film	Osc film	Xuong-Hamlin	Osc film	Xuong-Hamlin
Cell dimensions					
a, b, (Å)	60.9	60.9	61.2	60.9	61.0
c (Å)	96.8	96.8	95.2	96.9	95.8
Unique reflections	14,562	10,804	4,017	14,844	13,465
Resolution (Å)	1.75	1.8	2.7	1.7	1.9
Completeness of (data %)	65	54	82	63	80
R _{merge} (%)	4.6	6.1	7.3	6.1	4.5
Benzene bound		No	No	Yes	Yes
Refined model					
R(%)	14.8	Not refined	Not refined	15.2	16.2
Δ _{bond} (Å)	0.015			0.014	0.016
Δ _{angle} (°)	2.1			1.9	2.2

A crystal of T4 lysozyme, equilibrated with a standard mother liquor consisting of 2.3M phosphate, pH 6.7, 0.23M NaCl and 1 μl m⁻² 2-hydroxyethyl disulphide, was placed in a silanized (Sigmacoat, SIGMA) glass capillary. Small columns of mother liquor were left at each end of the capillary. In addition, a small column (~2 mm) of pure benzene was placed between the crystal and the mother liquor. The columns of mother liquor not only maintained humidity but prevented the benzene from dissolving the wax seal. The crystal was allowed to equilibrate at room temperature for 10–14 days as liquid (presumably benzene) condensed on the crystal but did not have any apparent effect on the diffraction quality of the crystal. Before X-ray exposure the capillary was opened, excess liquid removed from the crystal and the tube resealed. X-ray data were collected in part using oscillation photography (Osc film) and in part with a Xuong-Hamlin area detector system. R_{merge}, Agreement between repeated intensity measurements. R, Crystallographic residual for the refined model. Δ_{bond} and Δ_{angle} give the r.m.s. discrepancies between the bond lengths and angles in the refined model and their 'ideal' values. Coordinates will be deposited with the Brookhaven Data Bank.

largest backbone shifts (0.3–0.4 Å) occur in α-helix 108–113, which moves away from the bound benzene. The largest adjustments in side chains (up to 0.4–0.7 Å for Leu 84, Val 87 and Tyr 88) are for atoms in contact with or close to the benzene.

Leu 99 is completely buried in the hydrophobic core of T4 lysozyme and this is also true for the bound benzene. As noted, wild-type lysozyme has a cavity of volume 33 Å³ present in the carboxy-terminal domain adjacent to Leu 99 (Table 3). In the Leu 99 → Ala mutant structure, the volume of this cavity increases to about 150 Å³. The presence of the bound benzene reduces the cavity size to 32 Å³, essentially its original value. If the surrounding protein atoms were tightly packed around the side-chain of Leu 99, the Leu 99 → Ala replacement would not create sufficient space for a benzene. But, the geometrical arrangement is such that the benzene can be accommodated with very little change in the structure. If the benzene is placed, by model building, in the L99A structure at the same position it occupies

in the L99A-benzene complex, the two closest distances to surrounding atoms are 3.12 Å to C^{γ1} of Val 111 and 3.21 Å to C^{γ1} of Val 87. In the crystal structure of the L99A-benzene complex, these distances increase to 3.38 and 3.58 Å, respectively (Fig. 1d). In the complex the benzene is placed very symmetrically with respect to the β-carbon of Ala 99, with each atom essentially equidistant (3.6–3.9 Å) (Fig. 1d).

The crystallographic thermal factors for each of the six benzene carbons is relatively low, varying between 18–25 Å² in L99A and 26–37 Å² in the double mutant. This indicates that the benzene is bound in the crystal structure with high occupancy. But if the benzene ring is rotated azimuthally by 30° and refined in this alternative position, very similar thermal factors are obtained. This suggests that the axial angle of benzene is not well defined, and the benzene could be rotating relatively freely about the axis normal to the ring plane.

The benzene is buried in the hydrophobic core of T4 lysozyme with about 7 Å or more protein interposed between it and the surface. The admission of the ligand must require fluctuations in the protein structure of about 3.5 Å, the van der Waals' thickness of benzene. At the same time the structural changes can occur notwithstanding the constraints imposed by crystal lattice contacts. The fluctuations are therefore more consistent with the formation of 'mobile defects'¹⁰ than with opening and closing of the molecule as a whole. It provides further evidence of the dynamic nature of protein structures as previously shown by the 'flipping' of the internal side-chains^{11–13}, the penetration of proteins by external agents¹⁴, by hydrogen exchange¹⁵ and in various crystallographic contexts^{3,16–20}.

The site at which benzene binds is already preformed in the L99A and L99A/F153A mutant structures and the structural adjustments that occur on binding are minor. Under these circumstances the binding is favourable, as shown by the stabilization of the folded protein, and specific to one site. By contrast, the cavity formed by the Phe 153 → Ala substitution is too small to admit benzene without substantial conformational adjustments (about 1.5 Å) and there is no evidence that benzene tends to occupy this site. The results show that cavities can be engineered in proteins and suggest that such cavities can be tailored to bind a given ligand specifically. The purpose of such a binding site might simply be to sequester a desired ligand or as a basis on which chemical functionality might be added. The introduction of a hydrophobic cavity into a protein is expected to reduce stability but this might be offset by stabilizing amino-acid replacements engineered elsewhere in the protein. □

TABLE 3 Cavities in wild-type, mutant and benzene-containing lysozymes

Lysozyme	Cavity name	Area (Å ²)	Volume (Å ³)	Centre		
				X(Å)	Y(Å)	Z(Å)
WT*	I	52.4	33.5	28.0	8.9	0.6
	II	39.5	20.4	32.1	4.0	-4.9
L99A	I	182.3	149.5	27.3	6.8	3.3
	II	47.7	26.9	32.2	4.3	-4.8
L99A +Bz	I	51.7	32.2	27.8	9.1	0.4
	II	44.0	24.1	32.4	4.2	-4.9
F153A	I	151.9	100.7	30.2	6.3	0.7
	II	53.7	32.1	32.2	4.4	-4.6
L199A/F153A	I	271.2	221.6	28.8	5.9	2.4
	II	62.4	39.3	32.3	4.6	-4.6
L99A/F153A +Bz	I	50.7	30.2	27.7	8.8	0.3
	II	63.9	39.5	32.5	4.5	-4.5
	III	112.0	76.8	31.5	5.0	0.8

Cavities were calculated using the method of Connolly²⁴ with a probe radius of 1.2 Å. Cavities I and II are the largest cavities that exist in the structure of wild-type lysozyme. Both are in the vicinity of Leu 99 and Phe 153. Cavity III is at the position of Phe 153. In F153A and L99A/F153A this cavity is part of cavity I. When benzene binds to the double mutant it separates cavity III from cavity I.

Received 28 August; accepted 28 October 1991.

- Richards, F. M. A. *Rev. Biophys. Bioeng.* **6**, 151–176 (1977).
- Connolly, M. L. *Int. J. Peptide Prot. Res.* **28**, 360–363 (1986).
- Schoenborn, B. P., Watson, H. C. & Kendrew, J. C. *Nature* **207**, 28–30 (1965).
- Wishnia, A. *Biochemistry* **8**, 5070–5075 (1969).
- Eisenberg, H., Josephs, R. & Reiser, E. *Adv. Prot. Chem.* **30**, 101–181 (1976).
- Hermans, J. & Scheraga, H. A. *J. Am. Chem. Soc.* **83**, 3283–3292 (1961).
- Schellman, J. A. *Biopolymers* **14**, 999–1018 (1975).
- Tronrud, D. E., Ten Eyck, L. F., Matthews, B. W. *Acta crystallogr.* **A43**, 489–503 (1987).
- Dao-pin, S., Alber, T., Baase, W. A., Wozniak, J. A. & Matthews, B. W. *J. molec. Biol.* **221**, 647–667 (1991).
- Lumry, R. & Rosenberg, A. *Collin Int. CNRS* **246**, 53–62 (1975).
- Campbell, I. D., Dobson, C. M. & Williams, R. J. P. *Proc. R. Soc. B* **189**, 503–509 (1975).
- Wüthrich, K. & Wagner, G. *FEBS Lett.* **50**, 265–268 (1975).
- Snyder, G. H., Rowan III, R. & Sykes, B. D. *Biochemistry* **15**, 2275–2283 (1976).
- Lakowicz, J. R. & Weber, G. *Biochemistry* **12**, 4171–4179 (1973).
- Hvidt, A. & Nielsen, S. O. *Adv. Protein Chem.* **21**, 287–386 (1966).
- Tilton, R. F., Kuntz, I. D. Jr & Petsko, G. A. *Biochemistry* **23**, 2849–2857 (1984).
- Kretsinger, R. H., Watson, H. C. & Kendrew, J. C. *J. molec. Biol.* **31**, 305–314 (1968).
- Bennett, W. S. & Huber, R. *Crit. Rev. Biochem. Phys.* **15**, 291–384 (1984).
- Ringe, D. & Petsko, G. A. *Prog. Biophys. molec. Biol.* **45**, 197–235 (1985).
- Faber, H. R. & Matthews, B. W. *Nature* **348**, 263–266 (1990).
- Matsumura, M. & Matthews, B. W. *Science* **243**, 792–794 (1989).
- Becktel, W. J. & Schellman, J. A. *Biopolymers* **26**, 1859–1877 (1987).
- Dao-pin, S., Baase, W. A. & Matthews, B. W. *Proteins struct. funct. Genet.* **7**, 198–204 (1990).
- Connolly, M. L. *Science* **221**, 709–713 (1983).

ACKNOWLEDGEMENTS. We thank T. Alber and A. Morton for suggesting related experiments, J. A. Schellman and P. H. von Hippel for insights and Sheila Pepiot for technical assistance. A.E.E. acknowledges the support of the Swedish Natural Science Research Council. This work was also supported in part by the NIH and the Lucille P. Markey Charitable Trust.

Structure of the recA protein-ADP complex

Randall M. Story & Thomas A. Steitz

Department of Molecular Biophysics and Biochemistry, Department of Chemistry, and Howard Hughes Medical Institute at Yale University, New Haven, Connecticut 06511, USA

THE recA protein catalyses the ATP-driven homologous pairing and strand exchange of DNA molecules¹⁻³. It is an allosteric enzyme: the ATPase activity is DNA-dependent^{4,5}, and ATP-bound recA protein has a high affinity for DNA, whereas the ADP-bound form has a low affinity⁶. In the absence of ATP hydrolysis, recA protein can still promote homologous pairing, apparently through the formation of a triple-stranded intermediate^{1,7-9}. The exact role of ATP hydrolysis is not clear, but it presumably drives the triplex intermediate towards products^{1,9,10}. Here we determine the position of bound ADP diffused into the recA crystal. We show that only the phosphates are bound in the same way as in other NTPases containing the G/AXXXGKT/S motif. We propose that recA protein may change its conformation upon ATP hydrolysis in a manner analogous to one such protein, the p21 protein from the *ras* oncogene. A model is presented to account for the allosteric stimulation of DNA binding by ATP. The mechanism by which nucleoside triphosphate hydrolysis is coupled to the binding of another ligand in recA protein and p21 may be typical of the large class of NTPases containing this conserved motif.

Although recA crystals in the hexagonal P6₁ form were grown in the absence of any added nucleotide¹¹, ADP can be diffused into the crystal and the structure of the complex determined. A difference electron-density map (Fig. 1) allowed ADP to be readily fitted and its coordinates to be refined. ADP binds in a cleft at the carboxy end of the β -sheet and interacts with residues in three regions of the protein: helix D, the loop connecting β -strand 1 to helix C, and the loop between β -strand 8 and helix H. There are only minor changes in the protein structure produced by ADP binding to these crystals. Recently, crystals isomorphous to these have been grown in the presence of 1mM Mg-ADP (C. Brautigam, R.M.S. and T.A.S., unpublished results).

The adenine base stacks on Tyr 103; a preference for binding adenine rather than other bases is a result of interactions between its N6 and the carboxylate of Asp 100, and its N3 and the backbone NH of Gly 265. Tyr 264, which can be photolabelled with 8-azido-ATP (ref. 12), lies near the bound nucleotide, but makes no specific contacts (Fig. 1).

The α - and β -phosphates are bound by a loop connecting β -strand 1 and α -helix C whose amino-acid sequence, ⁶⁶GPESSGKT (in single-letter code), matches the well known 'motif A' consensus sequence for NTP binding (G/AXXXGKT/S, where X is any residue)¹³. This motif is found in many, but not all NTP-binding proteins (see for example, refs 14-16). Adenylyl kinase¹⁷, elongation factor Tu (EF-Tu)^{18,19} and *ras* p21 (ref. 20) are three NTP-binding proteins whose structures are known and whose nucleotide binding sites contain the 'motif A' consensus sequence. The positions of the nucleotides relative to one another and to the conserved peptide can be compared using the transformation obtained by superimposing the 12 α -carbons centred on the conserved sequence (Fig. 2a). The positions of the α - and β -phosphates are very similar, but the base and sugar positions of GDP in EF-Tu and in p21 are quite different from those of ADP bound to recA protein. Binding of the phosphates at the conserved loop is thus independent of other interactions between protein and nucleotide; such interactions can vary widely in these otherwise dissimilar proteins.

A comparison of the topologies of the β -sheets containing this phosphate-binding loop (Fig. 2b) shows that, except for

the more closely related EF-Tu and p21, they are markedly different in all of these proteins: the number of strands in the sheet, the polarity of the strands, and strikingly, even the connections of the elements around the phosphate-binding loop, all differ. It should be noted that none of these proteins has the classic Rossmann fold²¹ for this sheet (strand order, 6-5-4-1-2-3). Motif B' (an aspartate residue at the carboxy terminus of a β -strand)¹³ is also found in a different relative position—strand 4 in recA protein, strand 3 in the other examples (Fig. 2b, and see below). Therefore structure predictions assuming a conserved topology among otherwise unrelated proteins sharing these motifs^{22,23} should be treated with caution. These differences in the domain topology lead us to believe that the presence of a similar phosphate-binding loop in many proteins is unlikely to have been the result of divergent evolution of the whole domain from an ancestral α/β protein²⁴. Instead it could have arisen as a result of convergent evolution or by exon shuffling²⁵.

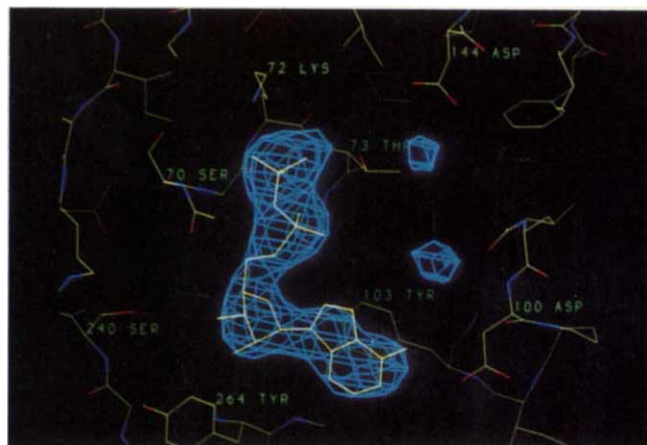
We expect that in addition to the α - and β -phosphates, the γ -phosphate of ATP bound to recA protein and of GTP to p21 will also be positioned similarly, as conserved residues in the phosphate-binding loop in p21 interact with the γ -phosphate and with a magnesium ion that bridges the β - and γ -phosphates²⁶. From model building, a plausible position for the γ -phosphate places it near the side chains of three residues outside the phosphate-binding loop: Asp 144, Glu 96 and Gln 194 (Fig. 3a). All of these residues are at the carboxy termini of strands of the β -sheet or in loop regions immediately following such strands.

Asp 144 and Glu 96 may have roles in catalysis (Fig. 3a). Aspartates at a position similar to Asp 144 and at the carboxy termini of β -strands (motif B),¹³ also occur in p21 (ref. 26), EF-Tu¹⁸ and adenylyl kinase²⁷ and are involved either directly or indirectly in binding the active-site magnesium ion. When Mn²⁺-ADP is soaked into these crystals, a difference electron density map indicates that the position of the bound Mn²⁺ is the same as that of Mg²⁺ in p21, EF-Tu and adenylyl kinase and that Asp 144 interacts with the Mn²⁺ ion through a water molecule (R.M.S., J. Tolman and T.A.S., unpublished results). Glu 96 is in a position to serve as a general base to activate a water molecule for an in-line attack on the γ -phosphate during ATP hydrolysis. A water molecule can be positioned at a distance of 3.4 Å from both the ϵ oxygen of Glu 96 and the phosphorus atom of the model-built γ -phosphate (Fig. 3a).

ATP and ADP stabilize different conformations of recA protein; the conformation stabilized by the binding of ATP has a higher affinity for DNA⁶. The close proximity of the known ATPase active site and the region likely to be involved in DNA binding²⁸ are suggestive of an allosteric mechanism by which the binding of ATP could affect the binding of DNA. Gln 194 (Fig. 3) is a good candidate for mediating a conformational change resulting from the binding of ATP. The carboxamide side chain of this residue bears roughly the same three-dimensional relationship to the γ -phosphate as does the peptide backbone of residues 59 and 60 of the 'switch 2' region of p21 (ref. 29), following the equivalent β -strand (Fig. 3b). In p21 the backbone nitrogens of these two residues interact with the γ -phosphate, leading to a large conformational change in the polypeptide region that follows—a poorly ordered loop and an α -helix²⁹. In recA protein, Gln 194 immediately precedes the disordered loop-2 region (residues 195-209) and helix G, which are likely to be involved in DNA binding²⁸. It is conceivable, then, that analogous to p21, this glutamine interacts with the γ -phosphate of ATP to stabilize a conformation of loop L2 and/or helix G that has a high affinity for DNA (Fig. 3c). Conversely, as the ATPase activity of recA protein is DNA-dependent, stabilization of such a conformation by DNA could be important in properly orienting Gln 194 or other residues in or near the active site, thereby enhancing ATP binding.

Many if not most of the proteins that contain the motif A consensus for nucleotide binding are proteins in which NTP

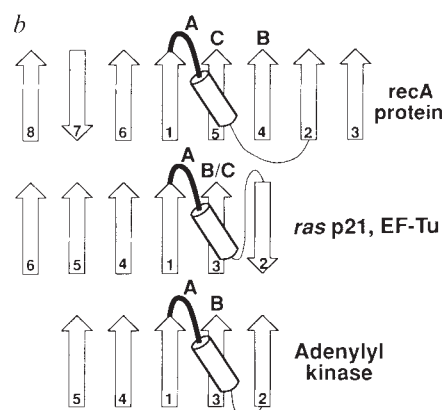
FIG. 1 Refined ADP model superimposed on difference electron density. The model of recA protein derived from a 2.3 Å resolution map of the protein in the absence of ADP²⁸ was refined using X-PLOR positional refinement versus data collected on a crystal soaked for 2 weeks in a solution containing 0.1 mM Mg²⁺-ADP. A difference electron density map at 2.7 Å resolution was calculated using observed (F_o) and calculated (F_c) diffraction amplitudes and calculated phases and contoured at 3.5 σ . This map allowed fitting of a model of ADP and its refinement. The R -factor was 28.4% before refinement, 23.0% after initial refinement and 22.4% after ADP was included in the model. Cell dimensions changed slightly to $a=102.2$ Å, $c=83.1$ Å, with only minor changes in the protein conformation. The two peaks to the right of ADP are water molecules that were omitted during the initial refinement and difference map calculation.



hydrolysis is coupled to the binding of another molecule (DNA, RNA or protein) through nucleotide-induced conformational changes. If we assume that such proteins also contain a phosphate-binding loop similar in structure to that of recA protein, there are constraints on how this coupling could occur. The properties of this loop provide tight binding for phosphates while positioning the γ -phosphate very near the carboxy end of a β -sheet. Such an arrangement leaves the γ -phosphate poised to interact with loops at the carboxy termini of β strands of globular α/β domains, leading to potential structural rearrangements. Such proteins can in general be considered as consisting of three regions A, B, and C, which are responsible for binding the three phosphates of the NTP (region A), Mg²⁺ binding and

ATP hydrolysis (region B), and the conformational change and perhaps target ligand binding (region C) (Fig. 3b, c). These designations are similar to those proposed by Walker *et al.*¹³ for the phosphate-binding loop (motif A) and the Mg²⁺ binding β -strand (motif B), but they include a third functionally common region which may not show any sequence similarity among these NTPases. Comparison of p21 and recA proteins shows that the restrictions imposed by motif B interacting with Mg²⁺ and by motif C interacting with the γ -phosphate allow these regions to lie on β -strands (or loops immediately following) that are either adjacent or one removed from the phosphate-binding loop, but no further away in the three dimensional structure. In the linear sequence, however, these motifs could be in order ABC or ACB,

FIG. 2 Comparison of recA protein with other proteins containing the phosphate-binding loop. *a*, Superposition of the α -carbon atoms of twelve residues centred on the G/AXXXGKT/S consensus sequence and corresponding nucleotide positions. Residues aligned are 64–75 in recA protein, 8–19 in p21²⁰, 16–27 in EF-Tu¹⁹, and 13–24 in adenyllyl kinase¹⁷. Root-mean square differences of the 12 α -carbons relative to recA protein are 0.40 Å (p21), 0.61 Å (EF-Tu), and 0.46 Å (adenyllyl kinase); r.m.s. differences between the positions of the 8 phosphate atoms (not included in the superposition) are 0.20 Å (p21) and 0.48 Å (EF-Tu). *b*, Comparison of the β -sheet topologies of recA protein, p21/EF-Tu and adenyllyl kinase showing significant differences. β -strands are depicted as arrows indicating their polarity from the amino to carboxyl end and are numbered according to their position in the primary sequence starting from the amino terminus. The phosphate-binding loop (A) in each case follows the first β -strand and is shown in bold. An α -helix, depicted as a cylinder, follows the loop. In the three classes shown the subsequent β -strand differs both in its spacing from the first β -strand and polarity in the sheet. Other connections are omitted for clarity. Motif B contains an aspartate involved in Mg²⁺ binding and located at the end of a β -strand (B) and is also found in different β -strands. A third region (C) is one of the two regions in p21 that undergoes a conformational change upon GTP binding via contacts with the γ -phosphate²⁹. The analogous region C of recA protein is proposed to be involved in ATP-coupled conformational changes and/or DNA binding as described in the text. RecA protein does not contain the 'core' structure proposed in ref. 24.



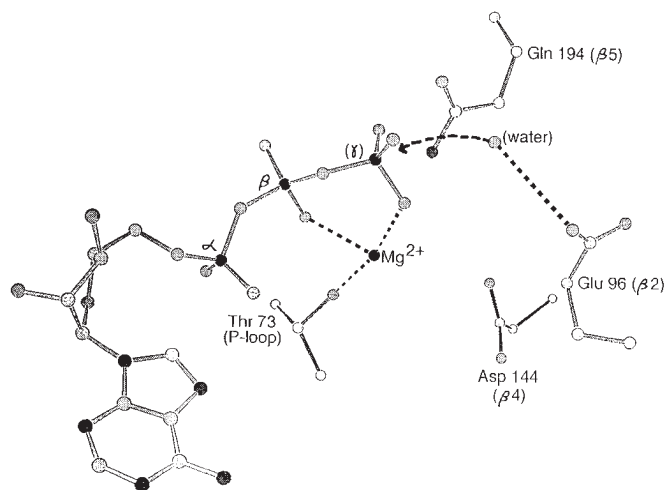
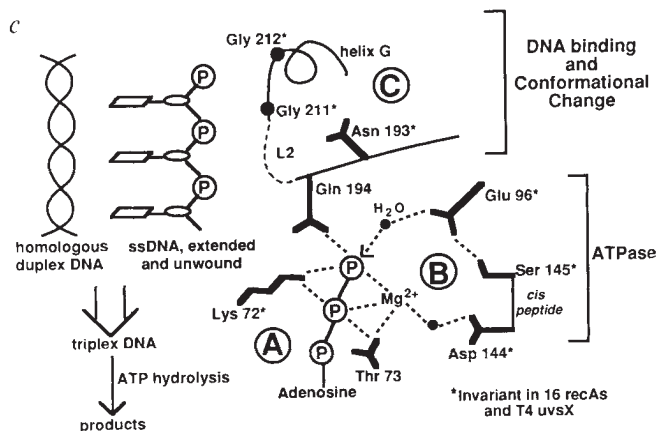
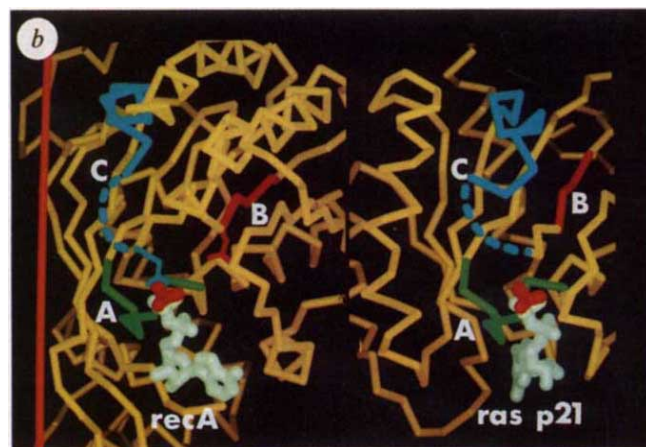


FIG. 3 Proposed catalytic and allosteric models for recA protein. **a**, Model of the active site, showing a possible enzymatic mechanism of ATPase involving the activation of a water molecule by Glu 96 for an in-line attack of the γ -phosphate. The positions of the γ -phosphate and water molecule are derived from model building. The position of the Mg^{2+} ion is based on that found in the GDP form of p21²⁰ and the alignment in Fig. 2a, as well as on a difference electron density map in which the Mg^{2+} was replaced by Mn^{2+} . The γ -phosphate was positioned assuming that Mg^{2+} bridges the oxygen atoms of the β - and γ -phosphates and that these phosphate oxygens, the Mg^{2+} ion and O γ of Thr 73 all lie in the same plane, as in p21 (ref. 26). The other atoms are those of the recA protein with ADP refined at 2.7 Å resolution. **b**, Comparison of the nucleotide-binding site of recA protein with that of p21. The two proteins have been similarly oriented by the superposition (Fig. 2a) of their phosphate-binding A motif structures (A, in green). The aspartic acid-containing B motifs (B) are coloured red with Asp 144 shown in recA protein. The partially disordered loop and following helix in the p21-GDP complex²⁰ (C), which changes conformation upon interaction with the γ -phosphate of GTP (ref. 29), is shown in blue. The analogously positioned disordered loop L2, helix G and Gln 194 in recA protein (C) is likewise in blue. The 6_1 axis along which the recA helical polymer packs²⁸ and near to which bound DNA must lie, is shown on the far left in red. Note that although much of the protein structure (in yellow) and the base and sugar of the NTP (in white) are different in these proteins, the relative positions of the α -, β - and presumably γ -phosphates (model built in red) and the A, B and C regions is very similar. **c**, Schematic model for the structural mechanism of allosteric interaction between ATP hydrolysis and DNA binding in recA protein. For conceptual purposes, regions of the protein around the active site are divided into three regions, A, B and C, as described in the text. All residues shown are invariant among all the bacterial recA proteins compiled in ref. 30 and in yeast dmc1 (D. Bishop and N. Kleckner, personal communication). Gln 194 and Thr 73 are conservatively substituted by His and Ser respectively in T4 uvsX (ref. 30). In this figure loop L2 and/or helix G are assumed to be important in DNA binding. Regions in or around loop L1 may also be involved in DNA binding, particularly of



homologous duplex DNA. The ATP-bound (high-affinity) state is depicted here. We propose that through its interaction with Gln 194, the γ -phosphate of ATP stabilizes loop L2 and/or helix G in a state that binds single-stranded DNA. The single-stranded DNA is held in such a way as to permit formation of a triple-stranded intermediate. ATP hydrolysis destroys interactions between the nucleotide and Gln 194, resulting in a change in conformation of the DNA-binding region(s), leading to generation of products. This model could also be extended to a four-stranded reaction, where primary binding of double-stranded DNA would occur, generating a four-stranded intermediate. This allosteric model may be extended to all of the ATPases and GTPases containing a phosphate-binding loop (region A). The binding site for the target macromolecule (protein, RNA or DNA) may be formed upon interaction between the γ -phosphate and a loop (region C) that lies adjacent to the phosphate-binding loop. The presence or absence of the γ -phosphate may directly alter the conformation and binding properties of region C in many NTPases.

depending on the topology of the β -sheet, and widely separated, but in the folded structure they are brought close together. Perhaps then, a key to understanding the function of such diverse NTPases as recA protein, myosin, membrane-associated transport proteins and RNA helicases is in part a matter of locating these regions in the linear sequence. □

Received 16 September; accepted 20 November 1991.

- Howard-Flanders, P., West, S. C. & Stasiak, A. *Nature* **309**, 215-220 (1984).
- Radding, C. M. *J. biol. Chem.* **266**, 5355-5358 (1991).
- Kowalczykowski, S. C. A. *Rev. Biophys. Chem.* **20**, 539-575 (1991).
- Ogawa, T. *et al. Cold Spring Harbor Symp. quant. Biol.* **43**, 909-915 (1979).
- Roberts, J. W., Roberts, C. W., Craig, N. L. & Phizicky, E. M. *Cold Spring Harbor Symp. quant. Biol.* **43**, 917-920 (1979).
- Menetski, J. P. & Kowalczykowski, S. C. *J. molec. Biol.* **181**, 281-295 (1985).
- Rao, B. J., Dutreix, M. & Radding, C. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2984-2988 (1991).
- Hsieh, P., Camerini-Otero, C. S. & Camerini-Otero, R. D. *Genes Dev.* **4**, 1951-1963 (1990).
- Menetski, J. P., Bear, D. G. & Kowalczykowski, S. C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 21-25 (1990).
- Rao, B. J., Jwang, B. & Radding, C. M. *J. molec. Biol.* **213**, 789-809 (1990).

- McKay, D. B., Steitz, T. A., Weber, I. T., West, S. C. & Howard-Flanders, P. *J. biol. Chem.* **255**, 6662 (1980).
- Knight, K. & McEntree, K. *J. biol. Chem.* **260**, 10185-10191 (1985).
- Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. *EMBO J.* **1**, 945-951 (1982).
- Higgins, C. F. *et al. Nature* **323**, 448-450 (1986).
- Linder, P. *et al. Nature* **337**, 121-122 (1989).
- Dever, T. E., Glynn, M. J. & Merrick, W. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1814-1818 (1987).
- Dreusicke, D., Karplus, P. A. & Schulz, G. E. *J. molec. Biol.* **199**, 359-371 (1988).
- Jurnak, F. *Science* **230**, 32-36 (1985).
- La Cour, T. F. M., Nyborg, J., Thirup, S. & Clark, B. F. C. *EMBO J.* **4**, 2385-2388 (1985).
- Tong, L., deVos, A. M., Milburn, M. V. & Kim, S.-H. *J. molec. Biol.* **221**, 751-754 (1991).
- Rossman, M. G., Moras, D. & Olsen, K. W. *Nature* **250**, 194-199 (1974).
- Mimura, C. S., Holbrook, S. P. & Ames, G. F.-L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 84-88 (1991).
- Hyde, S. C. *et al. Nature* **346**, 362-365 (1990).
- Milner-White, E. J., Coggins, J. R. & Anton, I. A. *J. molec. Biol.* **221**, 751-754 (1991).
- Gilbert, W. *Science* **228**, 823-824 (1985).
- Pai, E. F. *et al. EMBO J.* **9**, 2351-2359 (1990).
- Egner, U., Tomasselli, A. G. & Schultz, G. E. *J. molec. Biol.* **195**, 649-658 (1987).
- Story, R. M., Weber, I. T. & Steitz, T. A. *Nature* **355**, 318-325 (1992).
- Milburn, M. V. *et al. Science* **247**, 939-945 (1990).
- Roca, A. I. & Cox, M. M. *Crit. Rev. Biochem. molec. Biol.* **25**, 415-456 (1990).

ACKNOWLEDGEMENTS. This work was supported by an NIH grant to T.A.S. and an NSF predoctoral fellowship to R.M.S.